

Bachelorthesis

über das Thema

**Datenschutzrelevante Spuren in Messwerten intelligenter
Stromzähler**

Autor: David Hübner
Martikel Nr.: 22951

Prüfer: Prof. Dr. Joachim Charzinski
Dipl.-Ing. (FH) Christoph Lindenmüller

Eidesstattliche Erklärung

„Hiermit versichere ich, David Hübner, an Eides Statt, dass ich die vorliegende Bachelorarbeit mit dem Titel: "Datenschutzrelevante Spuren in Messwerten intelligenter Stromzähler" selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.

Ich habe die Bedeutung der eidesstattlichen Versicherung und die prüfungsrechtlichen Folgen (§26 Abs. 2 Bachelor-SPO (6 Semester), § 23 Abs. 2 Bachelor-SPO (7 Semester) bzw. § 19 Abs. 2 Master-SPO der HdM) sowie die strafrechtlichen Folgen (gem. § 156 StGB) einer unrichtigen oder unvollständigen eidesstattlichen Versicherung zur Kenntnis genommen.“

Ort, Datum

David Hübner

I Kurzfassung

In Smart Grids übermittelt der Stromzähler im Haushalt hochauflösende Messdaten des aktuellen Stromverbrauchs an den Energieversorger. Diese Daten ermöglichen es, sowohl dem Energieversorger, als auch potenziellen Angreifern, Aussagen über den Verbraucher zu treffen, die tief in die Privatsphäre eingreifen und Rückschlüsse auf die Lebensgewohnheiten erlauben. In dieser Arbeit wird die Aussagekraft dieser Informationen zunächst theoretisch erläutert und dann, am Beispiel eines Kühlschranks, überprüft. Ein wesentlicher Bestandteil des Vorgehens ist zunächst das automatische Erkennen und Extrahieren des Kühlschranks aus der Stromverbrauchskurve und die Analyse dieser Daten auf verschiedene Merkmale. Im zweiten Teil werden diese Merkmale anhand mehrerer Beispiele auf datenschutzrelevante Spuren hin untersucht. So lassen sich, trotz Unterschieden zwischen verschiedenen Kühlschrankmodellen, Anwesenheitsprofile der Benutzer erstellen und weitere Aussagen treffen.

II Abstract

In smart grids, smart meters transmit high resolution data of the current power consumption to the energy supplier. This data allows, both the energy supplier and potential attackers, to gather information about the consumer that intrudes into the privacy and allows to draw inferences about lifestyle habits. This thesis first shows the significance of this data in theory and then verifies the assumptions using the example of a fridge. An essential part of this approach is the automatic detection and extraction of the fridge from the load profile and the analysis of this data to detect certain characteristics. The second part then focuses on analysing these characteristics for privacy sensitive traces. As a result it is possible to create behavioural profiles of the consumers, despite differences between different devices.

III Inhaltsverzeichnis

I	Kurzfassung	I
II	Abstract	II
III	Inhaltsverzeichnis	III
IV	Abbildungsverzeichnis	V
V	Tabellenverzeichnis	VI
VI	Listing-Verzeichnis	VI
1	Einleitung	1
1.1	Motivation	1
1.2	Überblick über die Arbeit	1
2	Smart Grid und intelligente Stromzähler	2
2.1	Smart Grid	2
2.2	Smart Meter / Intelligente Stromzähler	2
2.3	Auswirkungen auf die Privatsphäre	4
3	Erkennung von Gerätesignaturen im Stromverbrauch	6
3.1	Motivation des Experiments	6
3.2	Erfassung der Messwerte	6
3.3	Kühlschrank als primäres Versuchsobjekt	7
3.4	Zustände eines Kühlschranks	9
3.5	Algorithmus zur Erkennung eines Kühlschranks	10
3.5.1	Erkennung der Startevents	10
3.5.2	Erkennung der Endevents	13
3.5.3	Parametrisierte Erkennung	14
3.5.4	Tool zur Umsetzung des Erkennungsalgorithmus	15
3.6	Entfernen des Kühlschranks aus den Messwerten	18
3.7	Überlagerung mit anderen Geräten	23
4	Auswirkungen auf die Privatsphäre	24
4.1	Vorüberlegungen zur Auswertung	24
4.1.1	Komplementäre Verteilungsfunktion der Intervalle	24
4.1.2	Kriterium für die Validierung	26
4.1.3	Beurteilung des Klassifikators	27
4.2	Auswertung verschiedener Beispiele	29
4.2.1	Skript zur automatischen Auswertung	29
4.2.2	Erste Analyse von Kühlschrank A	29
4.2.3	Betrachtung einer langen Phase ohne Aktivität	32
4.2.4	Nutzerverhalten über mehrere Kühlphasen	35
4.2.5	Anwendung der Pausenbetrachtung auf andere Modelle	39

4.2.6	Zusammenhang mit dem Energieverbrauch	41
5	Fazit	44
5.1	Zusammenfassung und Bewertung	44
5.2	Ausblick	45
6	Quellenverzeichnis	47

IV Abbildungsverzeichnis

Abb. 1	Stromverbrauchskurve Kühlschrank	8
Abb. 2	Zustände Kühlschrank (einfach)	9
Abb. 3	Startevent	10
Abb. 4	Endevent	13
Abb. 5	Parameterbestimmung	14
Abb. 6	Annäherung durch exp. Funktion	19
Abb. 7	Resultat der Entfernung	22
Abb. 8	Komplementäre Verteilungsfunktion von Kühlschrank A	24
Abb. 9	Lampenverteilung über 2 Tage	26
Abb. 10	Umgebung des false-positive Werts	30
Abb. 11	Dauer der Zustände ohne Aktivität	32
Abb. 12	Dauer der Zustände mit Aktivität	33
Abb. 13	Erkennung der Anwesenheit	37
Abb. 14	Pausen und Kühldauern bei Kühlschrank A	40
Abb. 15	Energieverbrauch/Dauer bei Kühlschrank B	41
Abb. 16	Energieverbrauch Vergleich	42

V Tabellenverzeichnis

Tab. 1	Messwerte eines weiteren Startevents von Kühlschrank A	12
Tab. 2	Ergebnis verschiedener Grenzwerte	29
Tab. 3	Ergebnis verschiedener Grenzwerte (bei Pausendauern vgl. Abb. 12)	39
Tab. 4	Ergebnis verschiedener Grenzwerte (bei Pausendauern vgl. Abb. 11)	39

VI Listing-Verzeichnis

Lst. 1	Erkennung Kühlschrankevents (als Pseudocode)	16
Lst. 2	Entfernen des Kühlschranks aus der Kurve (also Pseudocode) . . .	21

1 Einleitung

1.1 Motivation

Durch Smart Metering lässt sich der tatsächliche Stromverbrauch eines Haushalts im Sekundentakt exakt messen, um Haushalte in ein intelligentes Stromnetz (Smart Grid) einzubinden und Analysemöglichkeiten über das eigene Verhalten zu ermöglichen. Es liegt allerdings die Vermutung nahe, dass die Daten, die hier anfallen, auch für andere Zwecke verwendet werden können. Dies stellt auch der Bundesbeauftragte für Datenschutz fest und bemerkt: "Es gilt insbesondere zu vermeiden, dass Profile der Lebensführung von Menschen gebildet werden können. Dies käme einem Eingriff in die Privatsphäre entsprechend einem 'messtechnischen Lauschangriff' gleich." ([Kon12])

Es lassen sich ohne größeren Aufwand zahlreiche Szenarien finden, in denen diese Daten bedenklich sind. So reicht es bereits aus, sich die Ausmaße des NSA-Überwachungsskandals von 2013 anzuschauen, um zu sehen, wie und in welchem Umfang Daten gespeichert und im Zweifelsfall abseits ihres ursprünglichen Verwendungszweckes genutzt werden. Aber auch aus wirtschaftlicher Sicht sind Informationen über den Energieverbrauch eines Haushaltes oder einer Firma und die Schlüsse, die sich daraus ergeben, sehr interessant.

1.2 Überblick über die Arbeit

Im Rahmen dieser Arbeit soll nun untersucht werden, in welchem Ausmaß aus den Daten, die von intelligenten Stromzählern gesammelt und etwa an den Stromversorger übermittelt werden, Rückschlüsse auf Verhaltensmuster der Benutzer gezogen werden können, die in die Privatsphäre dieser eingreifen. So werden im ersten Schritt die Grundlagen zu den Themenkomplexen Smart Metering und Erfassung der Messwerte erklärt und dann darauf aufbauend im zweiten Schritt versucht, ein Anwesenheitsprofil der Nutzer zu erstellen. Die Arbeit beschäftigt sich in erster Linie mit der isolierten Betrachtung weniger Geräte und führt dann am Ende die Implikationen auf realistischere Beispiele aus.

2 Smart Grid und intelligente Stromzähler

2.1 Smart Grid

In den letzten Jahren hat sich die Ausgestaltung unseres Energienetzes deutlich geändert. Während noch vor einigen Jahren zentrale Stromerzeuger einen dominanten Anteil hatten, geht die jüngere Entwicklung eher in eine dezentrale Richtung mit vielen verschiedenen Stromerzeugern, bis hin zu kleinen Photovoltaikanlagen, die privat betrieben werden und überschüssige Energie direkt in das Netz einspeisen. Zusätzlich haben sich natürlich auch die Möglichkeiten zur Planung und automatischen Regulierung im Laufe der Jahre verbessert. Vor allem der Fall der Unterversorgung, also zu wenig verfügbarer Energie, um den erforderlichen Bedarf zu decken, stellt hierbei für den Stromversorger ein Problem dar und dieser muss dann oft kurzfristig teure Reserven nachkaufen. Eine möglichst präzise Vorhersage über den tatsächlich erforderlichen Verbrauch zu einem bestimmten Zeitpunkt ist hier also erforderlich. Dies wird, wenn man den Trend hin zu schlechter planbaren, erneuerbaren Energiequellen betrachtet, immer wichtiger. Ein Windkraftwerk kann nicht auf Bedarf Strom liefern, sondern ist in erster Linie von den Wetterverhältnissen abhängig.

Das Smart Grid stellt nun den Ansatz dar, alle Akteure innerhalb des Stromnetzes miteinander zu verbinden und die Kommunikation zwischen diesen zu ermöglichen, um diese erhöhte Planbarkeit zu erreichen.

2.2 Smart Meter / Intelligente Stromzähler

Die Komponente des Smart Grid, welche den Verbraucher direkt betrifft ist der intelligente Stromzähler. Dieser zeichnet, im Unterschied zu herkömmlichen Zählern, auch Ferraris-Zähler genannt, nicht nur die insgesamt konsumierte Energie auf, sondern erfasst deutlich detailliertere Daten über den Stromverbrauch. Die Auflösung, mit der hier gearbeitet wird, befindet sich im Bereich von wenigen Sekunden, womit vergleichsweise detaillierte Profile über das Nutzungsverhalten über einen Tag hinweg angelegt werden können. Es wird also der tatsächliche Energieverbrauch im Zusammenhang zu der Zeit aufgezeichnet. Dieser Smart Meter ist an das sogenannte Smart Meter Gateway angeschlossen, welches als Bindeglied zu den anderen Teilnehmern, also etwa dem Stromversorger, fungiert und die Kommunikation mit diesen steuert.

Zunächst hat ein Smart Meter nicht zwingend eine Verbindung zum gesamten Energienetz. Der eigentliche Sinn besteht aber nun natürlich aus der Anbindung der

einzelnen intelligenten Stromzähler, sei es bei Großkunden, was bereits seit einigen Jahren Standard ist oder bei Privatkunden, an das intelligente Stromnetz. Der Stromversorger kann dann per Fernzugriff auf den Zähler zugreifen bzw. der Zähler übermittelt automatisch in festgelegten Zeitabständen die gesammelten Daten. Damit soll eine bessere Lastenplanung- und Verteilung ermöglicht werden.

Gerade im Privatkundenbereich befindet sich die Einführung der digitalen Stromzähler erst in der Startphase und die meisten Haushalte verfügen noch nicht über ein solches Gerät bzw. die vorhandenen Smart Meter übermitteln die aufgezeichneten Daten nicht automatisch an den Versorger.

In der Novelle des Energiewirtschaftsgesetz (EnWG) von 2011 wird der gesetzliche Rahmen geschaffen, um insbesondere im Bereich der Privatkunden eine Umstellung auf intelligente Stromzähler zu erlauben. So sollen, angelehnt an eine EU Richtlinie¹, bis 2020 80% der Verbraucher über einen Smart Meter im Haushalt verfügen. Es sollen laut § 21c Abs. 1a EnWG unter anderem in Neubauten oder bei größeren Renovierungsarbeiten, "soweit dies technisch machbar und wirtschaftlich zumutbar ist", Messsysteme eingebaut werden, die sich als Smart Meter verstehen lassen.

Neben dem Nutzen für den Versorger, soll in der Theorie auch eine Optimierung des Stromverbrauchs für den Endkunden möglich sein. Vorstellbare Szenarien sind hier etwa die bessere Möglichkeit zur Verfolgung des eigenen Stromverbrauchs und daraus resultierende Möglichkeiten zur selbständigen Optimierung von diesem oder die Wahl eines entsprechenden Tarifs, der auf die Lastenverteilung über den Tag hinweg angepasst wird. Eine aktuelle Studie von Ernst & Young im Auftrag des Bundesministeriums für Wirtschaft und Technologie bescheinigt der flächendeckenden Einführung von intelligenten Strommessern in Deutschland in [Ern13] allerdings ein eher mageres Zeugnis. So fielen die durchschnittlichen Einsparungen durch die Einführung von Smart Metern bei privaten Haushalten mit einem Verbrauch von bis zu 6000 kWh mit 1,2-1,8% der Stromkosten (durchschnittlich 39 Euro pro Jahr) deutlich geringer aus, als die Anschaffungskosten, die etwa mit jährlich 89 Euro auf die betroffenen Kunden umgelegt werden könnten.

Wenn man nun zusätzlich die potentiellen Eingriffe in die Privatsphäre betrachtet, die noch vorgestellt werden, erscheint die Umstellung auf intelligente Stromzähler für den Endverbraucher nicht wirklich attraktiv.

¹Richtlinie 2006/32/EG des Europäischen Parlaments und des Rates vom 5. April 2006

2.3 Auswirkungen auf die Privatsphäre

Wie bereits erwähnt ergeben sich aus den vom Smart Meter aufgezeichneten und an den Stromversorger übermittelten Daten etliche datenschutzrechtliche Bedenken. Im folgenden sollen einige Grundzüge erläutert und mögliche Angriffsszenarien skizziert werden.

Grundsätzlich wird die Privatsphäre der Nutzer in Deutschland durch das Recht auf informationelle Selbstbestimmung geschützt. In diesem Fall lässt sich zusätzlich die verfassungsrechtlich garantierte Unverletzlichkeit der Wohnung (Art. 13 GG) anwenden.

Wenn es also möglich ist, einzelne Geräte oder Geräteklassen im Stromverbrauch der Haushalte zu identifizieren und die genauen zeitlichen Abläufe der Geräte zu extrahieren, ist es für den potentiellen Angreifer, sei dies der Stromversorger selbst, Hacker, die sich Zugriff auf die übermittelten oder gespeicherten Daten verschafft haben oder etwa, weiter gesponnen, Ermittlungsbehörden, ein Einfaches, tiefgreifende Informationen über die Lebensgewohnheiten- und Abläufe der Verbraucher zu gewinnen. Zunächst lassen sich natürlich Aussagen über die Zeiten, in denen ein Nutzer zu Hause war tätigen und etwa eine Urlaubsphase erkennen. Weiterhin lassen sich über bestimmte Kombinationen, beispielsweise dem zeitgleichen Betrieb zweier Geräte in normalerweise unterschiedlichen Zimmern, auch Aussagen über die Anzahl der anwesenden Bewohner treffen. Wenn man also beispielsweise den gleichzeitigen Betrieb eines Föns und des Küchenmixers aus den gewonnen Daten herauslesen könnte, wäre dies ein klares Indiz auf mindestens zwei anwesende Nutzer. Ein erkanntes Babyphon würde die Vermutung eines Kleinkindes im Haushalt nahelegen. Alle genannten Szenarien sind natürlich auch insbesondere für die Werbeindustrie sehr interessant und könnten, bei erlangtem Zugriff darauf, etwa für Datensammlungen oder gezielte Werbemaßnahmen genutzt werden.

Die Aussagekraft dieser Inhalte geht sogar soweit, dass es nicht nur möglich ist den Betrieb eines Fernsehers nachzuweisen, sondern auch den angeschauten Film über den Stromverbrauch über die Zeit zu identifizieren. Diese Tatsache hat eine Forschungsgruppe der FH Münster in [GJL12] bereits 2012 nachweisen können und dies zeigt deutlich, in welchem Umfang sich die Daten nutzen lassen können.

Ein großer Bestandteil der Frage nach der Aussagekraft stellt die Auflösung der erfassten Daten dar. In [Mü10] wird festgestellt, dass "zwischen einer Auflösung von einer Sekunde bis zu einer Auflösung von 15 Minuten kaum ein Unterschied wahrnehmbar ist" und "erst bei einer Auflösung von 60 Minuten wird ein deutlicher

Unterschied sichtbar". Die herkömmlichen Smart Meter haben aber die Fähigkeit die Daten eben im Takt von wenigen Sekunden zu erfassen und auch zu speichern.

Die datenschutzrechtlichen Bedenken sind seit Aufkommen von Smart Metern weltweit bzw. im Speziellen in Deutschland Teil der Debatte. Das Bundesamt für Sicherheit in der Informationstechnik (BSI) hat dazu ein Schutzprofil für das Smart Meter Gateway formuliert, welches in erster Linie folgende Sicherheitsziele verfolgt: Zunächst steht die Forderung "seitens der Hersteller Firewall-Mechanismen in das Gateway zu integrieren", sowie "nur Kommunikationsverbindungen von innen nach außen" zuzulassen. Darüber hinaus sollen die Daten "sämtliche[r] Kommunikationsflüsse [...] nach einer gegenseitigen Authentifizierung grundsätzlich verschlüsselt und integritätsgesichert" ([Bun14] S. 14) übertragen werden.

Auch wenn so zumindest grundlegende Sicherheitsstandards gewährleistet sind, lassen sich damit unberechtigte Zugriffe auf die Daten nicht prinzipiell ausschließen und es bleibt immer noch die große Frage, wie die Energieversorger mit den Daten umgehen, da selbst bei einer Ende-zu-Ende-Verschlüsselung vom Smart Meter Gateway beim Kunden bis zum Energieversorger keine Aussagen über die Verwendung der Daten bei Letzterem gemacht werden können.

3 Erkennung von Gerätesignaturen im Stromverbrauch

3.1 Motivation des Experiments

Zunächst ist es sinnvoll zu klären, welche Vorgehensweisen zur Untersuchung der Aussagekraft der Stromverbrauchskurven in Hinblick auf datenschutzrelevante Spuren überhaupt in Frage kommen. So könnte man etwa die Geräte theoretisch modellieren und versuchen die Auswirkungen auf deren Stromverbrauch so zu verstehen. Dieses Experiment soll allerdings in Form eines empirischen Beweises durch die tatsächliche Messung der Geräte mit einem Smart Meter durchgeführt werden, um so den gewonnenen Daten des Stromversorgers möglichst nahe zu kommen. Anhand dieser tatsächlichen Messungen wird dann versucht, die Aussagekraft über den Nutzer zu bewerten.

In einer realistischen Umgebung wird der Smart Meter den Stromverbrauch von etlichen Stromverbrauchern erfassen und die resultierende Stromverbrauchskurve wird damit eine Überlagerung von diversen Geräten sein. Zu Beginn des Experiments werden allerdings als Vereinfachung die Kurven der einzelnen Geräte isoliert betrachtet, um auf diesem niedrigen Level erste Ergebnisse zu erzielen.

3.2 Erfassung der Messwerte

Zur Erfassung kommt ein Raspberry Pi Minicomputer mit angeschlossenem "Easy Meter" als Stromzähler zum Einsatz. Diese Installation wurde so im Rahmen eines vorhergehenden Projekts festgelegt. Die höchste zu erzielende Genauigkeit betrifft dabei einen Messwert alle zwei Sekunden. Dieser Wert kommt aus dem Durchschnitt des Stromverbrauchs während dieses Intervalls zustande (vgl. [PS13]). Dadurch lassen sich, neben gewissen Ungenauigkeiten bei sehr kurzen Effekten unter Umständen auch einige Messungen erklären, die im Laufe des Experiments auftreten werden. Wichtig zu erwähnen ist in diesem Zusammenhang noch, dass der Stromzähler zwischen Steckdose und das gewünschte Gerät gesteckt wird und der Strom dann einmal durch den Smartmeter geleitet wird.

Der Easy Meter liefert die Daten dann an die serielle Schnittstelle des Raspberry Pi, wo sie von einem bereits vorhandenen Skript, das ebenfalls im Rahmen des HdM-Projekts entstanden ist, ausgelesen und gespeichert werden. Die Daten, die hierbei anfallen, sind neben dem aktuellen Timestamp, sowie der Systemzeit in lesbarem Format, einmal der aktuelle Verbrauch bzw. in diesem Fall eben der durchschnittliche Verbrauch in W über die letzten 2 Sekunden und die Gesamtenergie in kWh.

Letzterer Wert erlaubt später eine Betrachtung der insgesamt verbrauchten Energie über einen Zeitraum von gewissen Zeitschritten.

3.3 Kühlschrank als primäres Versuchsobjekt

Grundsätzlich lassen sich über viele Geräte Aussagen über das Nutzerverhalten treffen. Der erste Schritt besteht also darin, einen Verbraucher zu finden, der möglichst gut geeignet ist und damit eine hohe Aussagekraft besitzt. Hierfür muss ein Bewertungskriterium gefunden werden und es muss festgelegt werden, welche Eigenschaften ein Gerät erfüllen sollte, um diese Aussagekraft zu gewährleisten.

Mögliche Kriterien wären hierbei:

- Hohe Verbreitung
- Häufiges Vorkommen in einem Trace
- Vorhersagbarkeit
- Charakteristische Signatur des Verbrauchers
- Interaktion des Benutzers

Es macht also beispielsweise wenig Sinn ein Gerät zu wählen, das in unregelmäßigen oder vergleichsweise großen Abständen verwendet wird. Auch wenn etwa mit einer Waschmaschine durchaus Aussagen über den Nutzer getroffen werden können, beschränken sich diese häufig auf einen Tag in der Woche, der nicht zwingend in Zusammenhang mit den restlichen Tagen steht.

Ein Kühlschrank hingegen bietet eine ausreichende Erfüllung der Kriterien: er kommt in den meisten Haushalten vor, besitzt eine erkennbare Signatur, auch über verschiedene Modelle hinweg, die sich, wie sich noch zeigen wird, eigentlich nur durch gewisse Parameter unterscheidet und ist in regelmäßigen Abständen in dem Trace erkennbar. Weiterhin besteht eine direkte Korrelation zwischen Interaktionen mit dem Kühlschrank und Anwesenheit des (bzw. genauer: eines) Benutzers. Diese Interaktion hat normalerweise auch eine entsprechende Häufigkeit, so dass man folgende Annahme treffen kann:

”Wenn eine Interaktion mit dem Kühlschrank nachgewiesen werden konnte, war ein Benutzer zu Hause”

Die konkreten Fragen, die sich nun stellen, sind also zum einen die Festlegung, was man im Falle des Kühlschranks als Interaktion des Nutzers versteht und zum anderen, wie man diese anhand der Stromverbrauchskurve erkennen kann. Die Frage nach der Interaktion lässt sich recht einfach beantworten: es ist das Öffnen bzw.

Schließen der Kühlschranktür und der Vorgang zwischen diesen beiden Aktionen, also ggf. das Hereinlegen oder Herausnehmen von Waren. Allein das Öffnen des Kühlschranks bewirkt durch den Luftaustausch eine Erwärmung der Luft im Gerät. Dieser Effekt kann in manchen Fällen durch das Hereinlegen von warmen Waren zusätzlich verstärkt werden.

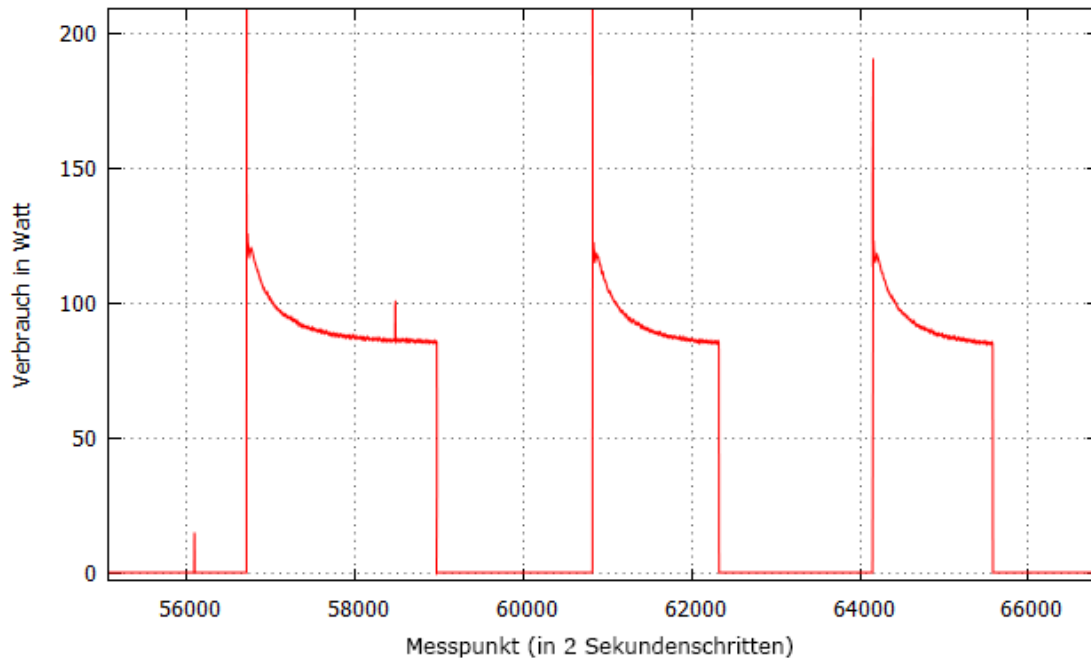


Abbildung 1: Typische Stromverbrauchskurve eines Kühlschranks

Abbildung 1 zeigt einen Ausschnitt der Messung eines Kühlschranks mit drei Kühlphasen. Zusätzlich lassen sich bei etwa 56000 und 58500 Messpunkten zwei kleine Peaks von etwa 15 Watt identifizieren, die sich durch die Lampe des Gerätes erklären lassen, welche beim Öffnen der Tür angeht. Dies wäre also eine erste Möglichkeit Benutzerinteraktion zu erkennen. Das Problem an dieser Methode ist allerdings die im Vergleich verschwindend geringe Leistung. Selbst wenn man nur den Kühlschrank betrachtet, ist diese gegenüber den anderen Werten sehr gering. Bei einer Überlagerung von diversen Verbrauchern wäre es nicht mehr möglich diese gezielt und zuverlässig zu identifizieren, da die Signatur *„kurzer Anstieg von etwa 15 Watt für wenige Sekunden“* nicht wirklich aussagekräftig ist. Die Lampen sind also nicht wirklich ein Weg, die Interaktion des Benutzers sicher zu erkennen, wohl aber - im Falle des isolierten Traces - eine Möglichkeit weitere Verfahren zu bewerten oder zu überprüfen.

Es liegt aber die Vermutung nahe, dass sich diese Form des äußeren Einflusses, also das Öffnen des Kühlschranks oder das Füllen mit neuen Waren, auch auf die

Dauer oder die Häufigkeit des Kühlprozesses niederschlagen wird und man über diese Daten, die sich in deutlich größeren Dimensionen bewegen, eine indirekte Aussage über die Interaktion treffen kann.

3.4 Zustände eines Kühlschranks

Wie in Abbildung 1 zu erkennen ist, besteht der Zyklus eines Kühlschranks in Bezug auf den Stromverbrauch, also im wesentlichen aus drei Zuständen. Auf eine Phase der Untätigkeit (idle), folgt ein kurzzeitiger starker Anstieg (peak) und danach ein exponentieller Abfall des Verbrauchs (der Kühlvorgang: cooling), bis eine bestimmte Grenze erreicht ist. Sobald diese Grenze unterschritten wird, erfolgt wieder der Übergang in den idle-Zustand. Stark vereinfacht lässt sich dieser Sachverhalt in folgendem Zustandsdiagramm darstellen:

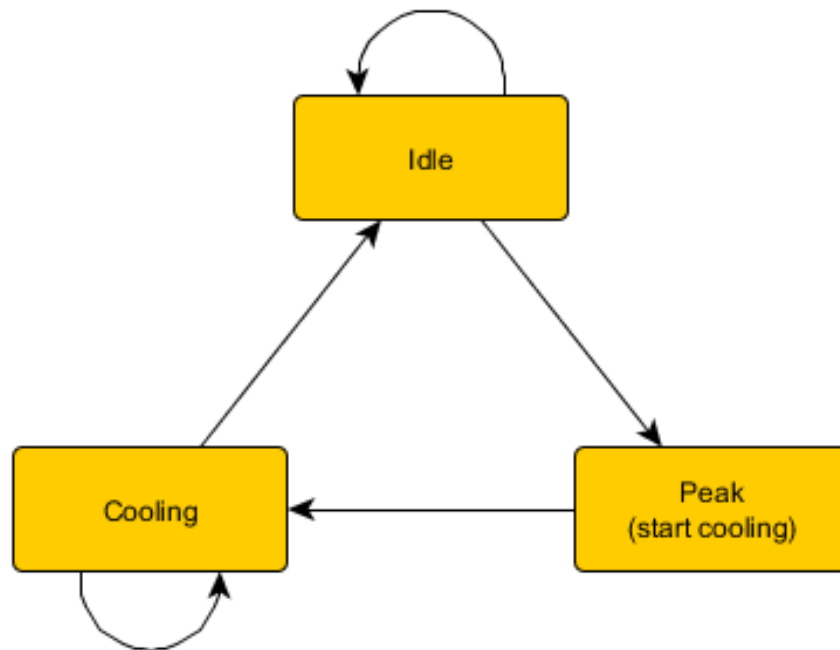


Abbildung 2: Einfaches Zustandsdiagramm Kühlschrank

Ziel soll in erster Linie das Erkennen bestimmter Events sein. Zunächst ist es hilfreich anhand der Zustände und derer Übergänge diese zu bestimmen. Im Kontext der Erkennung der Dauer einer Kühlphase (also der Dauer von cooling) oder der Pause zwischen zwei Kühlphasen (also der Dauer von idle), sind vor allem folgende zwei Events interessant, die am Beispiel von Abbildung 1 nachvollzogen werden können:

Beginn der Kühlphase (Startevent) Der Übergang vom Zustand idle zum Zustand

peak. Wie in Abb.1 ersichtlich ist, stellt dies in der Stromverbrauchskurve in erster Linie einen starken Anstieg von keinem Verbrauch auf rund 200 Watt dar.

Ende der Kühlphase (Endevent) Bei ungefähr 83 Watt endet die Kühlphase und der Zustand geht wieder in idle über.

Wenn sich diese beiden Events zuverlässig erkennen lassen, lässt sich über die beiden Intervalle [Beginn einer Kühlphase, Ende einer Kühlphase], also der im Zustand cooling verbrachten Zeit, sowie [Ende einer Kühlphase, Beginn der nächsten Kühlphase], was der Pause zwischen zwei Aktivitätsphasen des Geräts entspricht, die gewünschte Aussage konstruieren.

3.5 Algorithmus zur Erkennung eines Kühlschranks

Da nun klar ist, welche Vorgänge man erkennen möchte, lässt sich im folgenden Schritt ein Algorithmus formulieren, der die Beginn- bzw. Endevents anhand der Messdaten erkennen soll. Damit wird eine automatische Erkennung möglich, die auch bei längeren Messungen die Möglichkeit bietet, die Daten effizient analysieren zu können.

3.5.1 Erkennung der Startevents

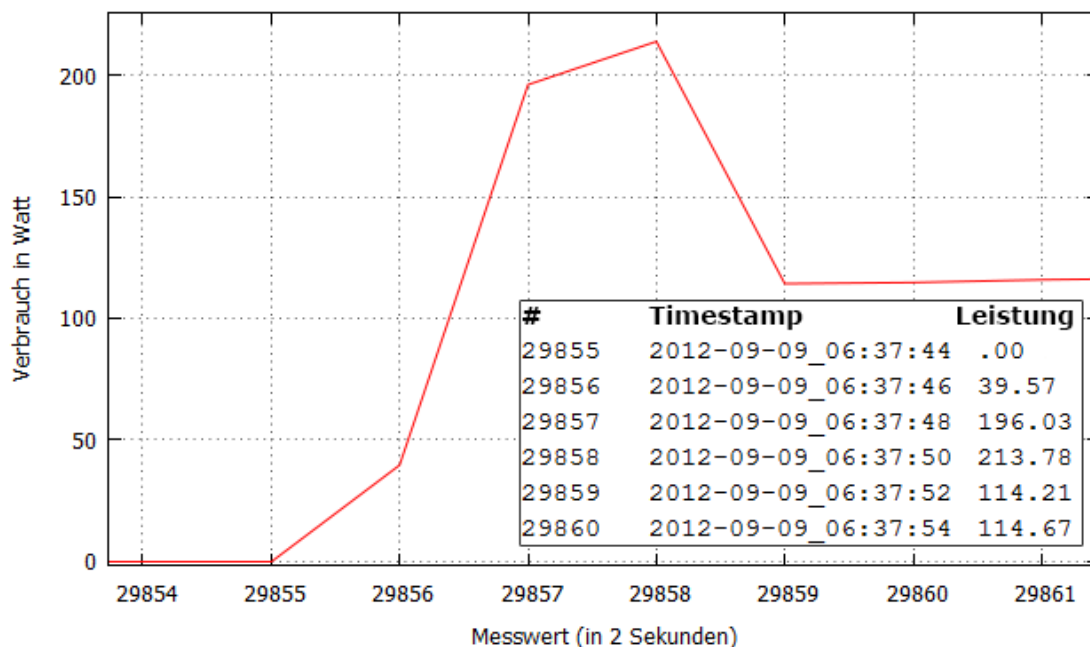


Abbildung 3: Beginn der Kühlphase (detailliert) mit Messwerten ¹

Zunächst ist es wichtig zu bemerken, dass die absoluten Werte keine Aussage erlauben, da diese in einer späteren Überlagerung mit anderen Geräten in anderen Größenordnungen liegen werden. Eine zuverlässige Erkennung lässt sich also nur über den Anstieg oder Abfall implementieren. Als weitere Vorüberlegung ist es wichtig festzustellen, dass der Erkennungsalgorithmus möglichst nicht zu speziell agieren soll. Eine Erkennung über die Abfolge von genau definierten relativen Veränderungen resultiert in zwei grundlegenden Problemen. Zum einen sind diese bei einem - isoliert betrachteten - Gerät durchaus unterschiedlich, was sich über die Messintervalle von 2 Sekunden und den unterschiedlichen Zustand des Kühlschranks zu Beginn der Kühlphasen erklären lässt. Dies lässt sich etwa an folgendem Beispiel anhand des darauffolgenden Kühlvorgangs verdeutlichen:

#	Timestamp	Leistung (W)
33035	2012-09-09_08:23:50	.00
33036	2012-09-09_08:23:52	137.34
33037	2012-09-09_08:23:54	234.33
33038	2012-09-09_08:23:56	154.00
33039	2012-09-09_08:23:58	113.96

Tabelle 1: Messwerte eines weiteren Startevents von Kühlschrank A

Auch wenn die grundsätzlichen Eigenschaften - Anstieg aus der idle-Phase von 0 Watt auf einen Peak von über 200 Watt und dann Übergang in den exponentiellen Abfall bei etwa 115 Watt - des Events ähnlich sind, lassen sich in den einzelnen Schritten doch starke Unterschiede feststellen. So ist hier beispielsweise nur ein Messwert zwischen Beginn der Phase und dem Peak, während sich dieser Anstieg in Abbildung 3 in zwei Messwerten niederschlägt. Zusätzlich sind die Störeinflüsse anderer Geräte bei einer Überlagerung als hoch einzuschätzen.

Auf der anderen Seite müssen noch weitere Kriterien verwendet werden, da sich ein Peak von einem bestimmten Wert nicht auf das Gerät "Kühlschrank" beschränken wird und eine solche Signatur im Falle einer Überlagerung vermutlich zu etlichen Fehlerkennungen führen würde. Nach diesen Vorüberlegungen lassen sich grob folgende Bedingungen formulieren, die charakteristisch für das Startevent einer Kühlphase des Kühlschranks sind:

¹Der Timestamp von 2012 kommt hier durch eine falsche Systemzeit des Raspberry Pi zu Stande und liegt somit außerhalb des Bearbeitungszeitraums dieser Arbeit. Gleiches gilt für die Angaben in Abbildung 4 und allgemein für die Messungen mit diesem Gerät ("Kühlschrank A").

Bedingung 1: Anstieg von über 160 Watt über einen Zeitraum von 3 Messwerten (6 Sekunden).

Bedingung 2: Zwischen zwei Startevents liegt ein gewisser Zeitraum. Bei dem hier betrachteten Gerät lässt sich beispielsweise 3000 Sekunden verwenden. Damit wird primär verhindert, dass innerhalb von einem Anstieg zwei aufeinanderfolgende Messwerte die Erkennung auslösen (im Beispiel von Abbildung 1 würde Bedingung 1 sowohl auf 196W, als auch auf 213W zutreffen). Ein weiterer Nebeneffekt ist, dass in einer Phase, in der definitiv keine weitere Kühlphase beginnen kann gar nicht die Möglichkeit einer Fehlerkennung besteht.

Bedingung 3: Direkt vor dem Peak ist die Leistung annähernd konstant.

Bedingung 4: Nach dem potentiellen Startevent fällt der Stromverbrauch ab (genaueres Modell: er fällt exponentiell ab). Hier ist es wichtig, den Abfall nicht nur anhand von einem Messwertpaar zu testen, da kurzfristige Störeinflüsse hier nicht auszuschließen sind (selbst wenn man nur den Kühlschrank betrachtet, könnte hier etwa die Lampe während der Kühlphase die Bedingung nicht erfüllen).

3.5.2 Erkennung der Endevents

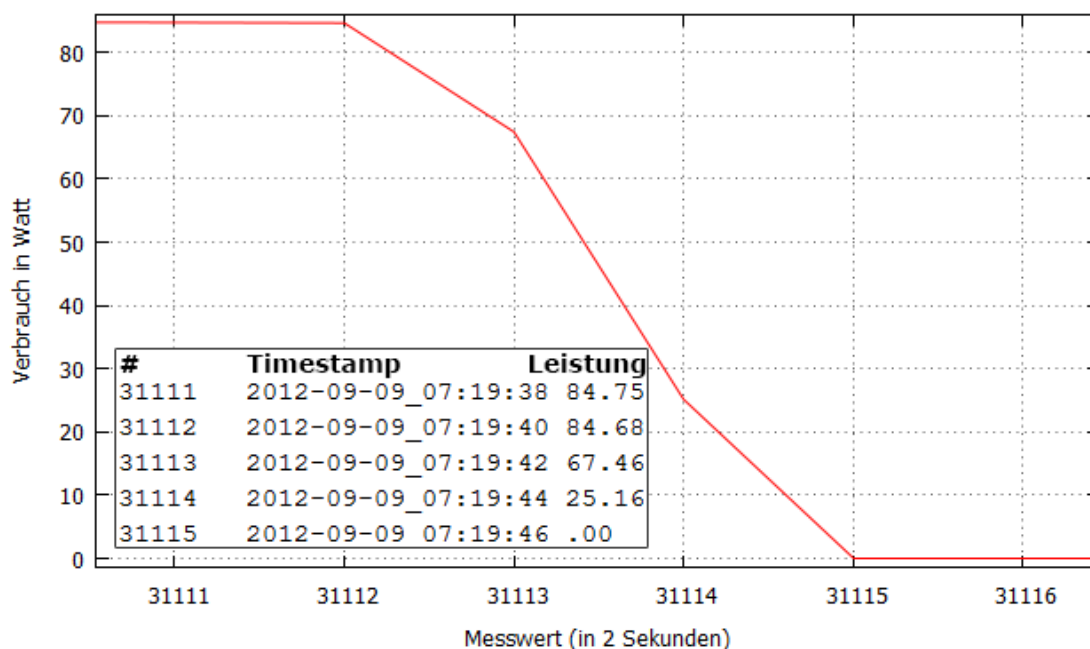


Abbildung 4: Ende der Kühlphase (detailliert) mit Messwerten

Analog zu den Überlegungen für die Erkennung des Beginns einer Kühlphase, lassen

sich auch für das Ende einer solchen gewisse Bedingungen formulieren, die eine gute Basis für die erfolgreiche Erkennung und möglichst wenig Fehler bieten:

Bedingung 1: Abfall der Leistung um mind. 75 Watt über einen Zeitraum von 3 Messwerten (6 Sekunden)

Bedingung 2: Verhindern der doppelten Erkennung, analog zum Startevent

Bedingung 3: Das Ende der Kühlphase kann nur auf ein Startevent folgen und hat einen Mindestabstand von 2000 Sekunden zu diesem. Die letzten beiden Bedingungen stellen sozusagen Plausibilitätsprüfungen dar, um den Zeitraum, in dem durch Einflüsse anderer Geräte Fehlerkennungen zustande kommen können, einzuschränken.

Bedingung 4: Vor dem Endevent befindet sich die Endphase des exponentiellen Abfalls und die Leistung ist annähernd konstant.

Bedingung 5: Gleiches gilt für die Phase nach dem Event.

3.5.3 Parametrisierte Erkennung

Wie in den Überlegungen zur Erkennung der beiden relevanten Ereignisse klar geworden ist, hängt der Erkennungsalgorithmus im Wesentlichen von einigen Parametern (z.B. von der Höhe der Änderung des Verbrauchs während des “Startpeaks“) ab, die sich von Gerät zu Gerät unterscheiden können. Nun ist es natürlich möglich, diese möglichst allgemein zu definieren, also einen großen Spielraum an möglichen Werten zuzulassen. Dieser generische Ansatz erlaubt eine akzeptable Erkennung über viele Kühlschranksmodelle hinweg, wird aber zu einer vergleichsweise höheren Fehlerrate, also in diesem Kontext false positive-Ergebnissen führen. Der Beginn des Heizvorgangs einer Kaffeemaschine etwa hat grob die gleiche Signatur in der Stromverbrauchskurve, zeichnet sich aber insbesondere durch einen klar größeren Stromverbrauch aus; der Peak beträgt rund 800 Watt. Wenn man nun den maximalen Wert des Peaks, den man noch dem Kühlschrank zuordnet, nicht beschränken würde, könnte dies aber trotzdem die Erkennung auslösen. Noch deutlicher wird dies im Falle der Erkennung der Endevents, da dies im Prinzip nur aus dem Abfall um einem bestimmten Wert besteht. Hier würde in vielen Fällen z.B. das Ausschalten einer Lampe ausreichen. Wenn man sich aber etwa den Stromverbrauch in Abbildung 1 anschaut, ist hier der Abfall quasi immer identisch und kann stark spezifiziert werden. Dieser Zusammenhang soll nun anhand der Messung eines weiteren Kühlschranks (Kühlschrank B) gezeigt werden.

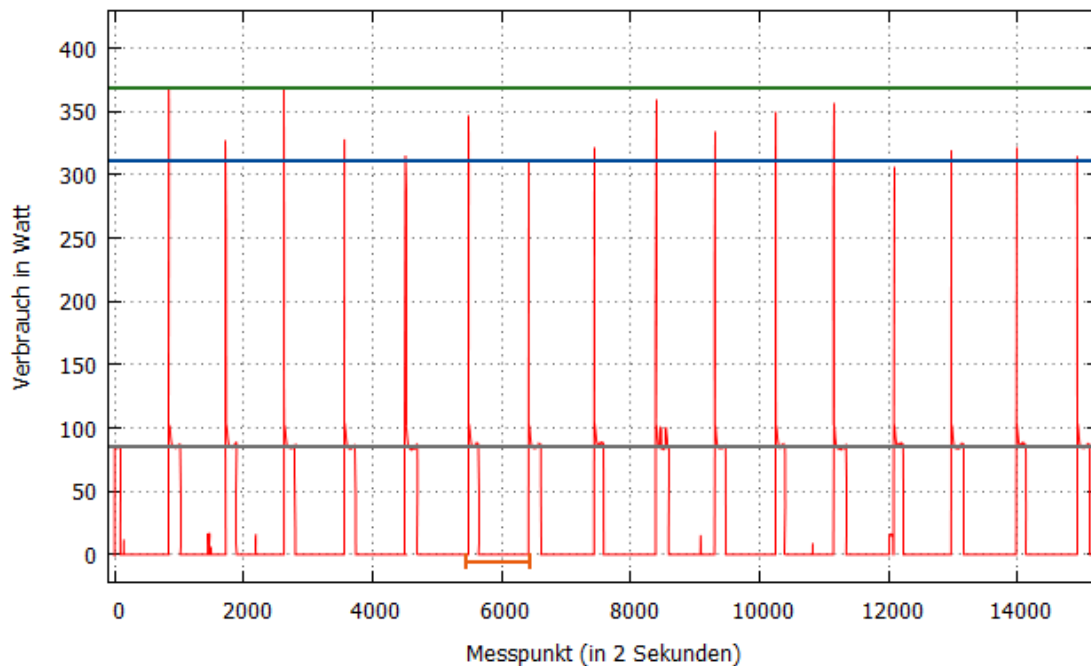


Abbildung 5: Stromverbrauchskurve Kühlschrank B

Es lassen sich hier zunächst folgende Parameter herauslesen:

PEAK_MAX_CHANGE Der Maximalwert des Ausschlags zu Beginn, hier etwa 370 Watt (grüne Linie).

PEAK_MIN_CHANGE Entsprechend das Minimum des Peaks, hier 310W (blau).

END_MIN_CHANGE Wie schon in der Betrachtung des vorherigen Geräts ist die Schwelle, der das Ende des Kühlvorgangs bedeutet, ziemlich exakt und beträgt hier etwa 86W (graue Linie). Es ist hier also ratsam sich in einem Bereich von $\pm 5\%$ zu bewegen, um etwaige geringfügige Abweichungen weiterhin zu erkennen. Als Minimalwert könnte man für dieses Gerät also 82W wählen...

END_MAX_CHANGE ...und als Maximalwert 90W.

MIN_GAP Der Minimale Abstand zwischen zwei Startevents liegt hier bei ungefähr 900 Messwerten (also 1800 Sekunden bzw. einer halben Stunde).

Nun stellt sich noch die Frage, wie man diese Parameter bestimmen kann. Der einfachste Weg ist, wie es hier auch erfolgt ist, die manuelle Festlegung über eine möglichst isolierte Messung des zu untersuchenden Gerätes. Dann lässt sich mit wenig Aufwand eine Festlegung der Parameter realisieren. Selbst in einer Messung mit mehreren Geräten wird sich an vielen Stellen noch recht problemlos ein Kühlschrank identifizieren lassen. Es ist davon auszugehen, dass nachts wenig Verbraucher aktiv

sind, während ein Kühlschrank auch zu dieser Zeit in regelmäßigen Abständen die Kühlphasen initiiert und die Höhe der Peaks bzw. der Abfälle bei den Endevents hat keine Korrelation mit etwaiger Interaktion des Benutzers. Ansonsten bleibt immer noch, wie bereits erwähnt, eine allgemeine Festlegung (z.B. Startpeak zwischen 150 und 400 Watt, Endabfall zwischen 75 und 95 Watt), die die meisten Gräte abdecken sollte.

3.5.4 Tool zur Umsetzung des Erkennungsalgorithmus

Im folgenden Schritt ist es nun also möglich einen Algorithmus zu formulieren, der diese Erkennung automatisch vornehmen kann. Im Rahmen dieses Experiments ist hierfür das Python Skript *fridgeAnalyze.py* zuständig, dessen Funktionsweise in diesem Kapitel skizziert werden soll. Der gewünschte Output ist eine Menge von Intervallen in der Form [Beginn der Kühlphase, Ende der Kühlphase], in dem der Kühlschrank gekühlt hat, sowie die daraus einfach abzuleitenden Kühldauern und Pausendauern. Diese Werte beziehen sich wieder auf die Messpunkte, also 2 Sekunden Schritte.

```

1  if ( (changeprev[0] >= PEAK_MIN.CHANGE or changeprev[1] >=
      PEAK_MIN.CHANGE or changeprev[2] >= PEAK_MIN.CHANGE) and
2  ((changeprev[0] < PEAK_MAX.CHANGE and changeprev[1] <
      PEAK_MAX.CHANGE and changeprev[2] < PEAK_MAX.CHANGE)):
3  if ( i-lastStart > MIN.GAP ):
4      if ( abs(getPower(i-3))-abs(getPower(i-2)) < 10 ):
5          if ( (abs(getPower(i+10))-abs(getPower(i+11)) < 15) or (abs(
              getPower(i+20))-abs(getPower(i+21)) < 15) ):
6              startEvent = true
7
8
9  if ( (changeprev[0] < END_MIN.CHANGE or changeprev[1] <
      END_MIN.CHANGE or changeprev[2] < END_MIN.CHANGE) and (not(
      changeprev[0] < END_MAX.CHANGE or changeprev[1] < END_MAX.CHANGE
      or changeprev[3] < END_MAX.CHANGE)) ):
10 if ( i-lastStart > MIN_IDLE.PHASE and lastStart > 0):
11 if ( i-lastEnd > MIN_IDLE.PHASE ):
12     check=1
13     for j in range(3,7):
14         if abs(getPower(i-j-1)-getPower(i-j)) > 20:
15             check=0
16     for j in range(3,7):
17         if abs(getPower(i+j)-getPower(i+j+1)) > 20:
18             check=0
19     if check == 1:
20         endEvent = true
21         addInterval(lastStart , i)

```

Listing 1: Erkennung Kühlschrankevents (als Pseudocode)

Die in Listing 1 skizzierte Funktionsweise des Erkennungsskripts setzt im Wesentlichen die theoretisch erarbeiteten Bedingungen um. Es wird die gesamte Liste an Messwerten durchlaufen und für jede Iteration wird als Vorbereitung ein Array generiert, das die Änderung der Leistung von dem aktuellen Zeitpunkt zu mehreren vorhergehenden Messpunkten beinhaltet. Diese Veränderungen werden nun in Zeilen 1 ff. bzw. 9 ff. zur Erkennung der Start- bzw. Endevents verwendet und stellen die am offensten formulierten Bedingungen dar. Wenn die Erkennung auf diese Eigenschaften triggert, wird der Zeitpunkt weiter auf Plausibilität geprüft. In den Zeilen 12 bis 18 wird im Kontext des Endevents schließlich geprüft, ob die Werte vor bzw. nach dem potentiellen Endevent annähernd konstant sind. Die in Zeile 21 dargestellte Funktion prüft dann zusätzlich noch, ob der Startzeitpunkt bereits in einem

anderen Intervall verwendet wurde, um zu verhindern, dass im Falle eines nicht erkannten Startevents ein fehlerhaftes Intervall generiert wird (Start der vorherigen Kühlphase bis Ende der aktuellen Kühlphase, wenn der Start der aktuellen Phase nicht erkannt wurde).

Hier wird auch deutlich, dass eine grundlegende Fehlerkontrolle bereits inhärent implementiert ist. Sollte ein Endevent nicht erkannt werden, überschreibt das nächste Startevent das bereits erkannte und nicht vervollständigte Startevent und die Länge der nächsten Kühlphase wird wieder korrekt erkannt. Nur wenn mehrere aufeinander folgende Events nicht korrekt erkannt werden, hat dies Auswirkungen auf die Intervalle. Hier wäre etwa eine weitere Plausibilitätsprüfung möglich, die die Intervalle auf eine maximal sinnvolle Dauer beschränkt und ansonsten verwirft.

3.6 Entfernen des Kühlschranks aus den Messwerten

Da nun ein Kühlschrank erfolgreich erkannt werden kann, ist es sinnvoll, sich darüber Gedanken zu machen, wie man im Falle von mehreren gemessenen Verbrauchern am besten vorgeht. Eine Möglichkeit ist das sequenzielle Verarbeiten der Stromverbrauchskurve, wobei in jedem Schritt ein Verbraucher identifiziert wird und dann dessen Signatur aus den Messwerten entfernt wird. Wenn also im Falle des Kühlschranks dieser möglichst gut modelliert werden kann, wird es möglich, in den durch den Erkennungsalgorithmus gefundenen Intervallen die Spuren dieses Geräts zu eliminieren. Damit ist es beispielsweise vorstellbar einfach zu erkennende bzw. zu modellierende Verbraucher (die dann vermutlich auch einen eher hohen Einzelverbrauch haben und die Gesamtkurve dominieren) früh zu verarbeiten, um nach und nach die Komplexität zu verringern, bis am Ende im Wesentlichen nur noch Rauschen, sowie die nicht erkannten oder geprüften Geräte vorhanden sind. Auch auf den Kühlschrank direkt bezogen ergibt sich durch das Vorgehen ein Anwendungsfall. So ist es möglich, den Kühlschrank an sich aus den Messwerten zu entfernen und nur die Lampe, welche später als Kontrollindikator für Anwesenheit des Nutzers herangezogen werden kann, zu behalten.

Wie in den vorherigen Kapiteln bereits festgestellt wurde, besteht die Signatur eines Kühlschranks in erster Linie aus drei Komponenten: Startpeak, Abfall zum Ende des Kühlvorgangs, sowie der Phase dazwischen, welche sich grob als exponentielle Funktion modellieren lässt. Die Höhe des Startpeaks lässt sich nicht eindeutig festlegen. Daher bietet es sich an, diesen über das Anpassen der Werte während den maximal 8 Sekunden des Peaks auf einen Wert kurz vor diesem Event zu eliminieren. Das Endevent muss im Grunde nicht weiter betrachtet werden, da man den exponentiellen Abfall einfach bis zu diesem Punkt weiter betrachten kann. Alternativ ist es analog zur Vorgehensweise bei dem Beginn der Kühlphase natürlich auch möglich in dieser Umgebung einen Wert aus der idle-Phase kurz nach dem Ende des Kühlvorgangs zu benutzen.

Die Aufgabe besteht nun also darin, den Abfall der Leistung während der Kühlphase möglichst gut zu modellieren. Wie bereits erwähnt lässt sich dieser offensichtlich über eine exponentielle Funktion annähern. Diese besitzt die Form:

$$f(x) = a + be^{cx}$$

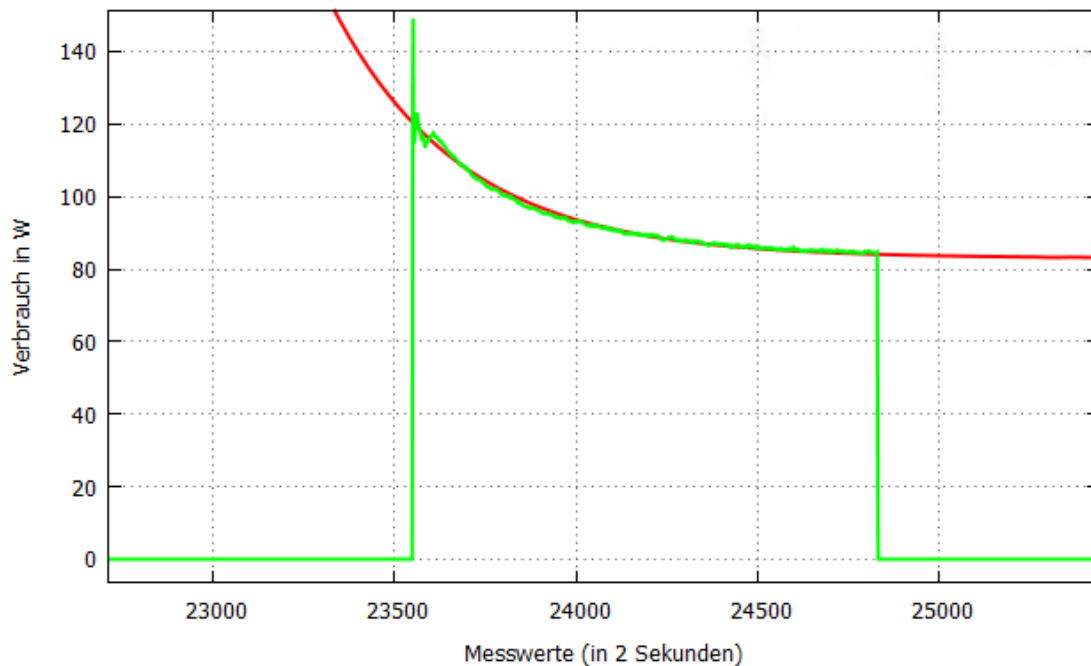


Abbildung 6: Annäherung durch exponentielle Funktion (rot)

Der genaue Verlauf der Funktion und damit die Werte der einzelnen Parameter unterscheiden sich jedoch von Fall zu Fall. Zum einen ist dies natürlich bei verschiedenen Kühlschränken der Fall, zum anderen aber auch abhängig von der Dauer des Kühlvorgangs. Der einfachste Ansatz, eine generische Belegung der Parameter zu wählen, ist also nicht wirklich zielführend. In diesem Zusammenhang wäre eine weitere Möglichkeit die feste Definition von verschiedenen Funktionen für gewisse Mindestauern der Kühlphasen. So ist im Allgemeinen bei einer kurzen Kühlphase der Abfall anfangs steiler, während bei einem längeren Intervall die Kurve eher flach erscheint. Im Falle des Kühlschranks aus Abbildung 1 wäre es also etwa möglich, drei Belegungen der Parameter zu finden, die charakteristisch für kurze Kühlphasen (etwa 2500 Sekunden, in Abbildung 1 die zweite und dritte dargestellte Phase), ausgedehnte Phasen (rund 3000 Sekunden), sowie lange Phasen (ab 3500 Sekunden, in Abbildung 1 das erste Beispiel) sind.

Ein Ansatz, der sich von dem bisher beschriebenen Vorgehen unterscheidet, ist die dynamische Bestimmung der Funktionsparameter über die vorliegenden Daten. Eine Option wäre der Versuch die Funktion über 4 Punkte aus der Kurve festzulegen. Dabei ergeben sich aber einige grundlegende Probleme. Im Falle einer Überlagerung mit anderen Verbrauchern ist es nicht gewährleistet, dass man wirklich Messwerte wählt, die auch auf dem störungsfreien Abfall liegen, was teilweise zu grob falschen Ergebnissen führen würde. Auch dies lässt sich wieder an einem einfachen Beispiel

anhand der Kühlschränklampen verdeutlichen. Würde man in der ersten Kühlphase aus Abbildung 1 zufällig einen Messpunkt während der Lampe bei knapp 59000 wählen, hätte die resultierende Annäherung eine erhebliche Abweichung vom eigentlichen Verlauf. Ein weiteres Problem ist die Tatsache, dass sich der Abfall eben nur grob über eine exponentielle Funktion modellieren lässt, einer solchen aber nicht unbedingt vollständig entspricht.

Sinnvoller erscheint das Fitten der Funktion in grob vorgegebenen Grenzen. Es ist offensichtlich, dass sich die Parameter in allen Fällen in einem gewissen Bereich befinden oder sich zumindest gut genug annähern lassen. Hierfür ist es zunächst hilfreich, sich ein Kriterium zu suchen, welches die einzelnen Iterationen des Fitting-Algorithmus bewertet. Naheliegender ist es, auf eine möglichst geringe Gesamtabweichung ϑ hin zu optimieren.

$$\vartheta = \frac{\sum_{i=1}^n \Delta x}{n}$$

Die Einzelabweichungen sind hierbei einfach der Unterschied zwischen den vorhandenen Messdaten und der exponentiellen Funktion aus der aktuellen Iteration. Um zum einen den Rechenaufwand gering zu halten und zum anderen den Einfluss anderer Geräte nicht zu stark werden zu lassen, ist es wie bereits erwähnt naheliegend, die Wertebereiche für die Parameter zu begrenzen.

a lässt sich gut eingrenzen, da dieser Wert letztendlich nur die Grenze angibt, ab der der Kühlvorgang beendet wird und das Gerät wieder in den idle-Zustand übergeht. Es bietet sich hier an, den in Kapitel 3.4.3 verwendeten Wert von `END_MIN_CHANGE` bzw. `END_MAX_CHANGE` wiederzuverwenden. Für b und c lassen sich auf Grund der zu erreichenden Steigung der Funktion sinnvolle Wertebereiche von 35 bis 40 bzw. -0.0022 bis -0.0030 festlegen. Bei signifikant anderen Kühlschränkmolellen ist es evtl. notwendig diese Parametergrenzen neu zu definieren.

Durch dieses Verfahren lassen sich in den meisten Fällen sehr gute Ergebnisse erzielen. Vor allem im Anfangsbereich beträgt die Abweichung zwar teilweise noch wenige Watt, was aber insgesamt nicht wirklich ins Gewicht fällt. Im folgenden soll die Implementierung dieses Verfahrens (*fridgeRemove.py*) demonstriert werden.

```

1  #Kuehlintervalle erkennen, siehe Listing 1
2  detectIntervals()
3
4  for interval in intervals:
5      #fitting
6      besta = 82
7      bestb = 34
8      bestc = 0.0027
9      bestdelta = 1000000
10     for a in range(82,88):
11         for b in range(34,39):
12             for c in range(0,7):
13                 thisc = 0.0027 + (c*0.0001)
14                 curdelta = 0
15                 for j in range(0,duration):
16                     old = values[interval_start+j]
17                     new = min(150,a+b *math.exp(-thisc*((j+interval_start)-
18                         interval_start+d)))
19                     delta = new-old
20                     curdelta += abs(delta)
21                     #wenn die aktuelle Wertekombination besser ist, uebernehmen
22                     if (curdelta < bestdelta):
23                         besta = a
24                         bestb = b
25                         bestc = thisc
26                         bestdelta = curdelta
27                 #waerend des Abfalls die gefundene E-Funktion abziehen
28                 for j in range(0,duration):
29                     old = values[interval_start+j]
30                     new = min(150,besta+bestb*math.exp(bestc*((j+interval_start)-
31                         interval_start)))
32                     delta = new-old
33                     interval_error += abs(delta)
34
35                 values[interval_start+j] = values[interval_start+j] - new
36             #Start und Ende entfernen
37             removeStartpeak()
38             removeEnd()

```

Listing 2: Entfernen des Kühlschranks aus der Kurve (also Pseudocode)

Wie bereits die Kommentare in Listing 2 verdeutlichen sollten, wird zunächst also der gesamte erlaubte Wertebereich durchprobiert und jeweils die Abweichung ϑ

über die gesamte Dauer hinweg berechnet. Die beste gefundene Kombination an Parametern wird dann ab Zeile 27 verwendet, um die Kühlphase aus den Messwerten zu eliminieren. Auf das Entfernen der Startevents wird hier nicht näher eingegangen, das besteht aber, wie bereits erläutert, im Wesentlichen aus dem Anpassen der Werte während dieses Vorgangs an einen "neutralen" Wert kurz vor dem Event. Das Ergebnis des Skripts lässt sich in folgendem Diagramm sehen:

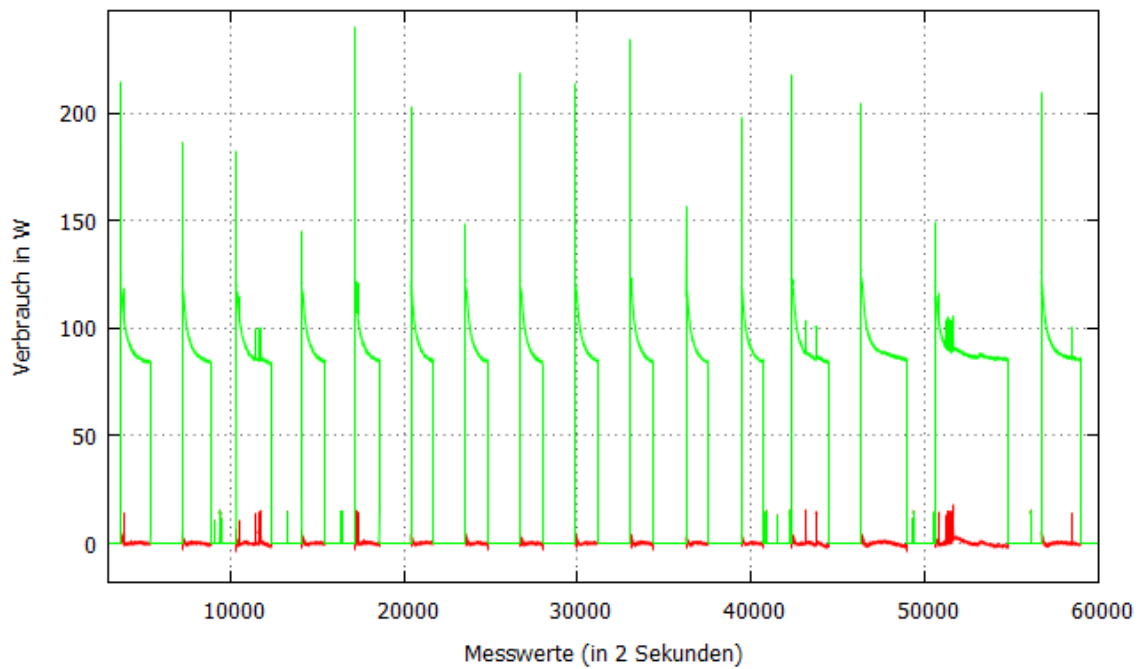


Abbildung 7: Ergebnis der Entfernung (rot), sowie Trace des Kühlschranks (grün)

Im Wesentlichen ist also nur noch Rauschen zwischen -3 und $+3$ W vorhanden und die Kühlschranklampen sind nun deutlich zu sehen. Die durchschnittliche Abweichung ϑ beträgt während den einzelnen Kühlphasen zwischen 0.5 W und 1.5 W, bei den längeren Phasen ab etwa 45000, die tendenziell etwas schwieriger zu modellieren sind. Hier ist natürlich auch klar, dass ϑ deutlich größer wird, je mehr Lampen oder andere Störeinflüsse innerhalb einer Kühlphase auftreten.

3.7 Überlagerung mit anderen Geräten

Nun geht es darum, zu überprüfen, wie gut der Erkennungsalgorithmus bei mehr als einem Gerät in der Stromverbrauchskurve funktioniert. Auf Grund des Aufbaus des Experiments bzw. der Art der Messung (es werden nur die Geräte erfasst, die direkt an den Smart Meter angeschlossen sind, nicht der gesamte Haushalt) wird diese Überlagerung von mehreren Geräten durch das Skript *addDevice.py* simuliert. Hier soll zunächst die grobe Funktionsweise erläutert werden und im Anschluss werden die Resultate diskutiert.

Da bisher das Hauptaugenmerk auf einem Kühlschrank lag, wird der Stromverbrauch eines solchen als Basis genommen und es werden nach und nach gewisse Geräte auf die Messwerte aufaddiert. Im Normalfall besitzen die meisten Geräte entweder Phasen der Aktivität (also beim Kühlschrank etwa die Kühlphasen) und Phasen der Inaktivität (der Idle-Zustand). Von dem Gerät, das man hinzufügen möchte, liegt nun eine gewisse Anzahl solcher Aktivitätsphasen in kurzer zeitlicher Abfolge vor. Man könnte eine Kaffeemaschine zum Beispiel einfach fünfmal in Folge brühen lassen, wobei jeder dieser Vorgänge wenige Minuten dauert und nur kurze Pausen zwischen den einzelnen Messungen liegen. Diese Pausen entsprechen dann den Inaktivitätsphasen, in denen der Verbrauch 0 Watt beträgt. Das Skript sorgt nun dafür, dass in diesen (manuell definierten) 0 Watt-Phasen zufällig verteilte Pausen eingefügt werden. So lassen sich die einzelnen Gerätesignaturen auf die zeitlich deutlich längere Messreihe des Kühlschranks verteilen. Es geht hier ja nicht darum, aus den Messwerten der anderen Geräte direkte Schlüsse z.B. auf das Nutzerverhalten zu ziehen (dann wäre eine solche zufällige Verteilung nicht zielführend), sondern nur darum, die Erkennung zu verifizieren.

Grundsätzlich funktioniert die Erkennung mit wenigen addierten Geräten noch gut. Hier ist jedoch auch zu beachten, dass es auf die Position bzw. die Phase, in der sich der Kühlschrank aktuell befindet, ankommt, wie gut die Erkennung noch funktioniert. Würde der Startpeak der Kaffeemaschine beispielsweise direkt auf den Startpeak der Kühlphase des Kühlschranks fallen, könnte dies zu einer nicht mehr erfolgreichen Erkennung führen.

4 Auswirkungen auf die Privatsphäre

Während es im letzten Kapitel in erster Linie um das Erkennen eines Gerätes ging, sollen nun im zweiten Teil des Experiments die gewonnenen Ergebnisse genutzt werden, um der eigentlichen Frage nach den Auswirkungen auf die Privatsphäre des Benutzers nachzugehen. Konkret geht es also zunächst darum, anhand der Kühlphasen und deren Dauer Rückschlüsse auf die Anwesenheit und das Verhalten des Nutzers zu ziehen.

4.1 Vorüberlegungen zur Auswertung

4.1.1 Komplementäre Verteilungsfunktion der Intervalle

Als ersten Schritt ist es angebracht, sich einen Überblick über die Dauer der einzelnen Phasen zu machen. Zu diesem Zweck wird die Funktionalität des Skripts *fridgeAnalyze.py* genutzt, das die Dauern der Kühlphasen erfasst hat. Anhand dieser Werte wird nun eine komplementäre Verteilungsfunktion generiert. Das Ziel dieses Vorgehens ist es, eine Grenze zu finden, ab der man eine Kühlphase als "lang" klassifizieren würde.

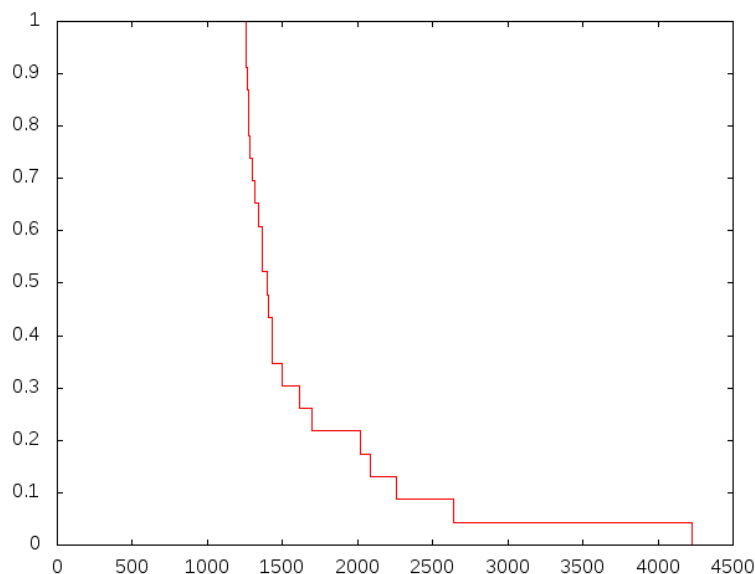


Abbildung 8: Verteilungsfunktion der Intervalle

In Abbildung 8 ist diese Verteilungsfunktion anhand einer Messreihe von 23 Kühlphasen in insgesamt knapp 2 Tagen von Kühlschrank A dargestellt. Jede Stufe entspricht einem Messwert, die Gesamtverteilung von 1 ist dabei also in Abschnitte von $\frac{1}{23}$ auf der y-Achse unterteilt. Die x-Achse stellt die Dauer der Intervalle dar. Es lassen sich

also beispielsweise folgende Aussagen treffen. 100% der Werte haben eine Dauer von 1250 oder mehr, dies ist also die Mindestdauer der Kühlphasen in der Messreihe. 30% der Werte betragen 1500 oder mehr und schlussendlich befindet sich nur einer der 23 Werte über 4200.

Es stellt sich die Frage, an welchem Punkt man eine Grenze ziehen möchte, ab der die Intervalle ungewöhnlich lang sind. Unterhalb dieser Grenze fiele die Abweichung der Werte dann in die Kategorie "natürliche Abweichung" durch Kühlschranksinterne Gründe oder etwa unterschiedliche Außentemperatur und oberhalb der Grenze könnte man größere äußere Einflüsse, also primär Interaktion des Benutzers und damit durch die Luftzirkulation bedingtes Aufwärmen der Luft in dem Kühlschrank vermuten. Um diese Grenze festzulegen gibt es prinzipiell mehrere Möglichkeiten, im folgenden werden drei davon aufgeführt, die dann auch im weiteren Verlauf des Experiments bewertet werden.

Einen Wert annehmen Hier stellt sich die Frage, wie oft man normalerweise überhaupt eine mögliche Interaktion des Benutzers und damit eine ungewöhnlich hohe Dauer erwarten würde. Die Annahme, dass alle Intervalldauern, die sich in der oberen Hälfte befinden, eine solche Einwirkung vermuten lassen, impliziert beispielsweise die Voraussetzung, dass der Nutzer die Hälfte der Zeit potentiell zu Hause und wach ist. Wenn man davon ausgeht, dass spät Abends und nachts die Aktivität gegen Null gehen sollte und tagsüber eine längere Abwesenheitsphase wegen Arbeit existiert, ist dies also eher nicht gegeben und einfach den Median der Wertemenge zu benutzen erscheint als nicht sinnvoll. Hier spielt auch die Tatsache hinein, dass man das Öffnen der Kühlschranktür in erster Linie um Essenszeitpunkte herum erwarten würde.

Den ersten größeren Knick verwenden In dem Beispiel aus Abbildung 8 befindet sich bei etwa 30% ein Knick in der Verteilung bzw. anders formuliert nimmt die Dichte der Messwerte pro Abschnitt auf der x-Achse ab. Man könnte hier also einen Übergang von normalen zu ungewöhnlich langen Messwerten erwarten. Zusätzlich erscheint die Annahme von grob einem Drittel anhand der Überlegungen aus dem letzten Punkt durchaus plausibel.

Anfangsabfall fortsetzen Mit diesem Verfahren wird der Abfall der ersten Phase, die man noch als normale Abweichung interpretieren kann (hier also etwa bis zu 60%) bis zur x-Achse fortgesetzt und dann orthogonal zu dieser wieder auf die Kurve gelegt. Damit kommt man zu einem Wert von etwa 25%, was im vorliegenden Beispiel etwa 1600 betragen würde.

Aus den bisher erfolgten Betrachtungen erscheint eine Wahl zwischen 25 und 30% also als sinnvoll. Dies bedeutet, übertragen auf die Intervalle also, dass man die Grenze zwischen 1500 und 1600 festlegen könnte. Um sich die Bedeutung dieser Werte nochmal klar zu machen: das sind die Messschritte im Abstand von jeweils 2 Sekunden. Eine Grenze von 1500 führt also zu der Annahme, dass eine Kühldauer von unter 3000 Sekunden, oder 50 Minuten, bei diesem Gerät normal ist.

4.1.2 Kriterium für die Validierung

Um die Annahmen zu überprüfen ist eine Möglichkeit ein Protokoll über die Anwesenheit eines Benutzers, dies war im Rahmen des Experiments aber nicht umsetzbar und außerdem lassen sich so oder so nur Aussagen über den Zusammenhang des Öffnen der Kühlschranktüre mit der Kühldauer treffen. Es reicht in diesem Fall also aus, das Öffnen der Türe, welches eindeutig durch der Stromverbrauch der Lampen aus einer isolierten Messung herausgelesen werden kann, als Kriterium für die Validierung zu verwenden.

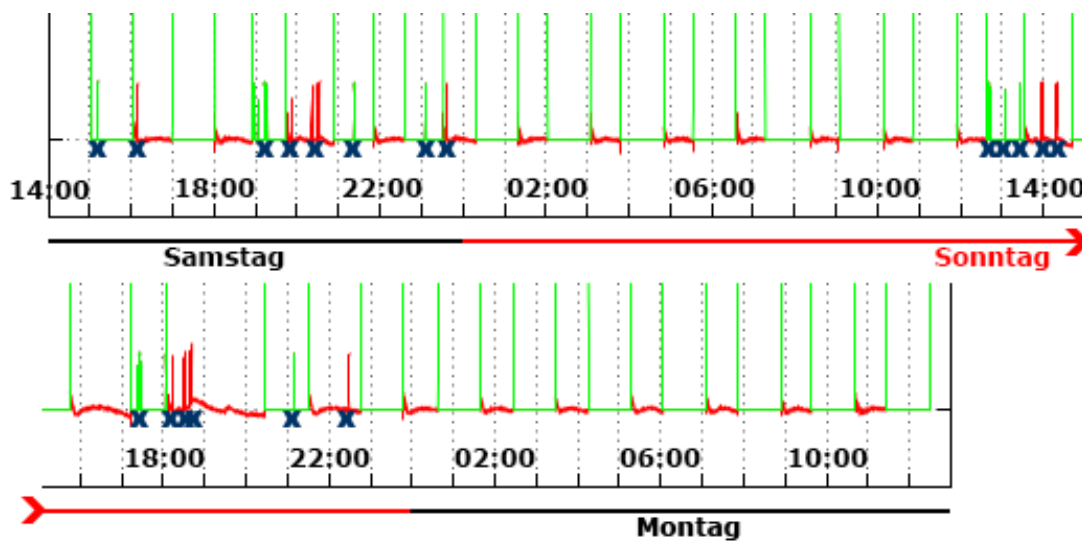


Abbildung 9: Verteilung der Lampenevents (blaue Kreuze) über zwei Tage

Abbildung 9 zeigt eine solche Verteilung der Lampen anhand der roten Kurve der entfernten Kühlphasen, sowie den 15W-Peaks aus der grünen Kurve während Idle-Phasen. Zur Verdeutlichung sind zusätzlich die Messpunkte in die entsprechenden Uhrzeiten übersetzt. In diesem Beispiel bezieht sich die Darstellung primär auf die beiden Tage Samstag und Sonntag. Man sieht sehr gut, dass sich an beiden Tagen drei Cluster bilden. Jeweils zur Mittagsphase um 14 Uhr, abends um bzw. nach 18 Uhr, sowie nochmal um etwa 0 Uhr (samstags) bzw. 22 Uhr (sonntags). Mit

diesen Werten lassen sich also ganz eindeutig Verhaltensmuster der Nutzer feststellen und Aussagen über die Anwesenheit um die genannten Cluster tätigen. Für die kommende Analyse der Vermutungen sollten die Indikatoren für das Öffnen der Kühlschranktür in einer leicht zu verarbeitenden Form vorliegen, also etwa als Liste der Zeitpunkte der Lampe. Diese kann entweder manuell erstellt werden oder über ein extra Skript anhand der Messwerte extrahiert werden.

4.1.3 Beurteilung des Klassifikators

Ziel der Analyse soll es nun sein, die einzelnen Kühlphasen in die beiden Klassen "normale Dauer - keine Interaktion" und "lange Dauer - vermutete Interaktion" einzuteilen und dann diese Einteilung zu beurteilen. Hierfür lassen sich folgende vier Kriterien verwenden.

true positive t_p

true negative t_n

false positive f_p

false negative f_n

True positive und true negative sind dabei korrekt erkannte Phasen, in denen respektive Nutzerinteraktion bzw. keine Interaktion, also Öffnen der Tür, nachgewiesen werden kann und false positive und false negative die entsprechenden falsch erkannten Pendants dazu. Interessant ist in erster Linie die Rate der richtig erkannten Phasen, also:

$$P(\text{richtigKlassifiziert}) = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \quad (1)$$

Es stellt sich nun die Frage, wie man die entsprechenden Kriterien definiert. Genauer ausgedrückt geht es darum, nach welchen Maßstäben einer erfolgten Klassifikation in "positive" bzw. "negative" das Attribut "true" bzw. "false" zugeordnet werden soll. Hierzu muss man sich überlegen, wann das Öffnen der Kühlschranktür vermutlich Einfluss auf die Länge einer Kühlphase haben wird. Wie bereits erläutert, ist das triviale Modell, welches diesem Vorgang zugrunde liegt der Austausch der gekühlten Luft im Kühlschrank mit wärmerer Umgebungsluft sowie evt. ein längerer Effekt im Inneren des Geräts durch das Hineinstellen von warmen Waren. In diesem Zusammenhang sollte auch die Möglichkeit betrachtet werden, dass der Sensor im Kühlschrank etwas träge auf die Temperaturänderung reagiert. Dies hat bei einem Gerät mit Kühldauern von durchschnittlich 40 Minuten natürlich einen geringeren

Einfluss auf die einzelnen Intervalle, wie bei einem Gerät mit Kühlphasen von wenigen Minuten.

Der naheliegende Ansatz ist hier die Festlegung, dass ein als positiv erkanntes Intervall, dann tatsächlich positiv ist und richtig erkannt wurde, wenn in dem Zeitraum zwischen Ende der vorherigen Kühlphase und Ende der aktuell betrachteten Kühlphase die Türe geöffnet wurde. Dies bedeutet also, dass man davon ausgehen würde, dass ein Erwärmen der Innentemperatur in der Idlephase bzw. der derzeitigen Kühlphase einen Einfluss auf die Dauer hätte. Ein etwas anderer Ansatz stellt das Verschieben dieser Grenzen nach vorne dar, also die Betrachtung des Zeitraums zwischen Beginn der letzten Kühlphase und Beginn der Aktuellen. Die Überlegung, die hier zugrunde liegt, ist die Trägheit des Sensors, der evt. nicht mehr in der aktuellen Kühlphase auf eine leichte Erwärmung reagiert, wenn diese gegen Ende auftritt. Hier sieht man auch die Möglichkeit, die Grenze beliebig zwischen diesen beiden Extremen zu verschieben, also etwa die Mittelpunkte der einzelnen Intervalle zu benutzen.

4.2 Auswertung verschiedener Beispiele

4.2.1 Skript zur automatischen Auswertung

Aufbauend auf die Überlegungen aus vorherigem Kapitel soll jetzt die Auswertung der gesammelten Kühlschrankschrankmesswerte erfolgen. Da eine manuelle Kontrolle der einzelnen Kühlphasen bei entsprechende Länge der Messungen aufwändig wäre, kommt auch hier wieder ein Skript zum Einsatz, welches als Hilfsmittel zu diesem Zweck dient. *fridgeValidate.py* benötigt als Input zwei Dateien, die im bisherigen Teil des Experiments bereits generiert wurden: zum einen die aus dem Erkennungsalgorithmus erzeugte Liste an Kühlintervallen und zum anderen eine Liste von Zeitpunkten, an denen die Interaktion des Nutzers nachgewiesen werden konnte (etwa die Lampenevents).

Abhängig von der Wahl der bereits diskutierten Definition der Einteilung in die beiden Klassen positive bzw. negative wertet das Skript nun einfach die gefundenen Intervalle aus und prüft ob in dem gewählten Zeitraum vor bzw. während der derzeit betrachteten Kühlphase ein Lampenevent auftrat und damit eine Einwirkung des Nutzers stattgefunden hat.

4.2.2 Erste Analyse von Kühlschrank A

Nach den Vorüberlegungen und der Erklärung der Funktionsweise von *fridgeValidate.py* ist es nun also möglich, als ersten Versuch das Beispiel der Verteilungsfunktion aus Abbildung 8 (Kühlschrank A) zu überprüfen.

Grenze	%	t_p	t_n	f_p	f_n	$P(\text{richtig Klassifiziert})$
1450	50	5	13	3	2	0.78
1480	40	5	13	3	2	0.78
1500	35	5	13	3	2	0.78
1580	30	5	14	2	2	0.83
1620	25	5	15	1	2	0.87
2050	20	3	15	1	4	0.78

Tabelle 2: Ergebnis verschiedener Grenzwerte

Die Annahme ist hier, dass der Kühlschrank länger kühlen muss, wenn im Zeitraum davon die Türe geöffnet wurde und sich dadurch die Innentemperatur erhöht hat. False positive f_p bedeutet in diesem Kontext also, dass die Kühldauer überdurchschnittlich lang war und dabei die festgelegte Grenze überschritten hat, aber kein Lampenevent nachgewiesen werden konnte, während false negative f_n eine Kühldauer unter der

Grenze mit allerdings nachgewiesener Interaktion bedeutet. Analog dazu sind die jeweiligen richtig erkannten Phasen zu interpretieren.

Das beste Ergebnis lässt sich mit einer Grenze von etwa 1620 Zeitschritten erreichen, was in Abbildung 8 den oberen 25% der Messwerte entspricht. Grundsätzlich kann man sagen, dass mit steigender Grenze die Anzahl der falsch als positiv erkannten Fälle (f_p) abnimmt, während die tatsächlich positiven (f_n) zunehmen. Dies lässt sich einfach erklären. Ein f_p -Fall ist in der Regel ein Intervall mit leicht überdurchschnittlicher Länge, die aber nicht durch äußere Einwirkung zustand kommt, aber bei entsprechend niedriger Grenze in die falsche Kategorie fällt. Je höher man nun diese Grenze legt, desto mehr Spielraum bleibt für solche Kandidaten. Ein Intervall, das als false negative (f_n) erkannt wird, zeichnet sich häufig durch eine leicht erhöhte Intervalldauer aus, die dann aber bei Anheben der Grenze recht schnell nicht mehr ausreicht, um eine richtige Einordnung vorzunehmen.

Interessant ist hier auch, dass sich das beste Ergebnis relativ nah am Durchschnittswerte der Einzeldauern von 1651 Messpunkten bewegt. Im folgenden sollen nun versucht werden, die drei falsch erkannten Werte zu erklären. Die beiden als false-negative erkannten Intervalle bewegen sich im Längenbereich von 1340 bzw. 1400, was im Vergleich zu den darauf folgenden Intervallen von etwa 1250 während der Nachtphase immer noch leicht erhöht ist. Hier ist es also evt. sinnvoll, nicht nur eine Grenze festzulegen, sondern eher eine lokale Untersuchung mit einfließen zu lassen. Der false-positive-Wert besitzt eine vergleichsweise sehr hohe Dauer von 2637 Messpunkten, hat aber im zur Zeit gewählten Betrachtungszeitraum von Ende der letzten Kühlphase bis Ende der aktuellen Kühlphase keine Lampenevents.

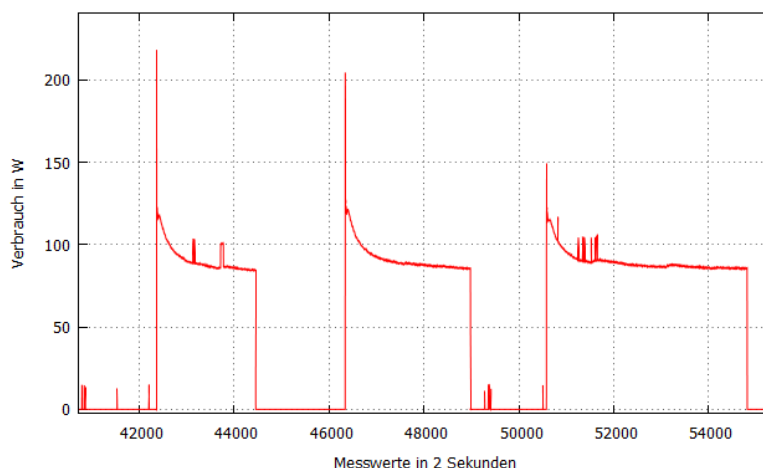


Abbildung 10: Umgebung des false-positive Werts (mittlere Kühlphase)

Auffällig ist hier natürlich, dass in vorherigen Kühlphase recht viele Interaktionen

des Benutzers registriert wurde und weiterhin, dass die darauffolgende Kühlphase die mit Abstand längste in dieser Messreihe ist. Die Vermutung, dass hier die Effekte der ersten dargestellten Phase also noch verzögerte Auswirkungen auf das falsch erkannte Intervall haben, liegt nahe. Dies könnte zum einen damit erklärt werden, dass das Öffnen der Tür während der ersten Kühlphase nicht direkten Einfluss auf die Dauer hat oder zum anderen mit der Möglichkeit, dass etwas Warmen in den Kühlschrank gelegt wurde, was auch in der mittleren Phase aus Abbildung 10 noch zusätzlich gekühlt werden muss. In diesem Zusammenhang ist es nun auch sinnvoll, die bereits angedeutete alternative Festlegung des zu betrachtenden Zeitraums für die Lampenevents zu analysieren. Wenn man nur die drei dargestellten Kühlphasen betrachten würde (die anhand der Dauer ja alle als positiv klassifiziert werden), würde die Bedingung "Nutzerinteraktion hat eine Auswirkung auf die Phase, wenn diese im Zeitraum zwischen Start der vorherigen und Start der aktuellen erfolgt ist" zu einem komplett richtigen Ergebnis führen. Wendet man diese Bedingung allerdings auf den gesamten Verlauf an, fallen wiederum andere Kandidaten in die falschen Kategorien, was insgesamt nicht zu einem besseren Resultat führt. Diese Änderung der Vorgehensweise ist also nicht zielführend und eine gewisse Rate an falsch zugeordneten Intervallen lässt sich hier also nicht vermeiden, da schlichtweg nicht alle Variablen, die dafür vermutlich notwendig wären, bekannt sind. Es ist davon auszugehen, dass neben den bereits erwähnten Faktoren von verzögertem Einfluss oder Kühlen von warmen Speisen über mehrere Intervalle hinweg, auch Punkte wie die Außentemperatur oder gerätespezifische Faktoren eine Rolle spielen.

4.2.3 Betrachtung einer langen Phase ohne Aktivität

Ein Ergebnis der Analyse des vorherigen Beispiels ist die relativ konstante Dauer der Kühlintervalle während den beiden längeren Phasen ohne Interaktion, welche sich naheliegenderweise auf die Nachtstunden zwischen 0 und etwa 7 Uhr festlegen lassen. Dies soll nun anhand einer langen Messung über knapp eine Woche überprüft werden, in der folgende Bedingungen gelten. Während der gesamten Dauer ist keine Interaktion eines Benutzers erfolgt, das bedeutet die Türe wurde nie geöffnet. Weiterhin ist die Außentemperatur relativ konstant und bewegt sich zwischen 17 und 18 Grad Celsius. Zusätzlich wurde als Vergleich noch eine weitere Messung mit dem gleichen Gerät durchgeführt, welche sich auf einen Zeitraum mit hoher Aktivität (rund um die Weihnachtsfeiertage) bezieht. Wenn man die gewonnen Erkenntnisse der Auswertung des ersten Beispiels heranzieht, würde man nun vermuten, dass die Intervalle in der ersten Messung annähernd gleich lang sind, während man aus der zweiten Messung wieder ähnliche Schlüsse auf das Nutzerverhalten wie im ersten Beispiel ziehen kann.

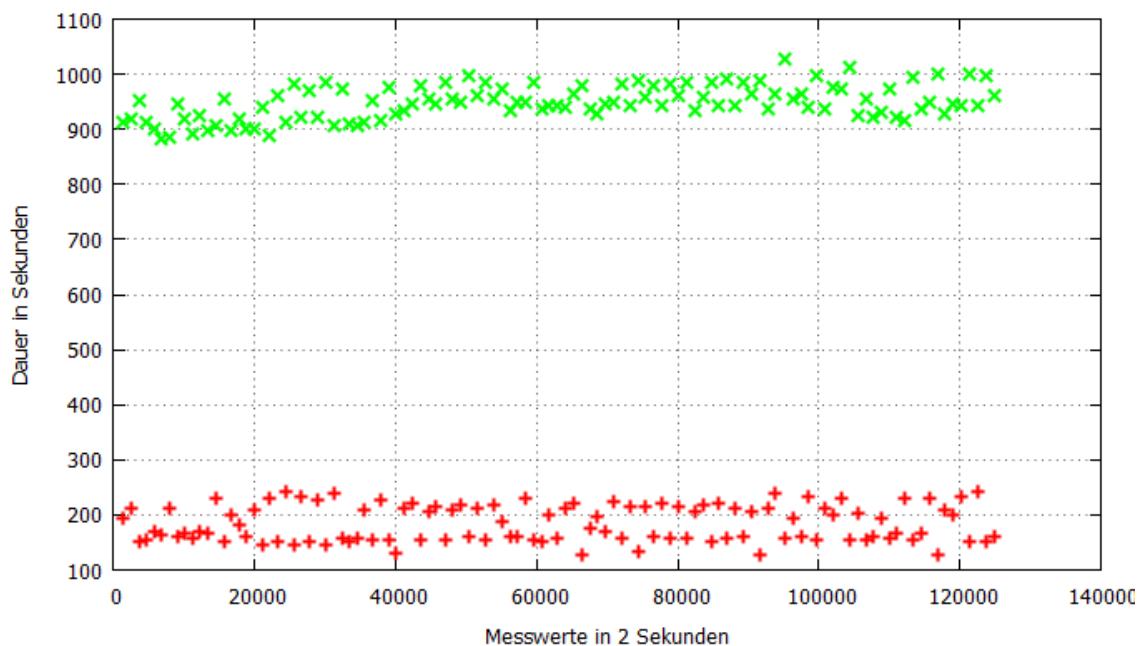


Abbildung 11: Dauer der Kühlphasen (rot) und Pausen dazwischen (grün) ohne Interaktionen

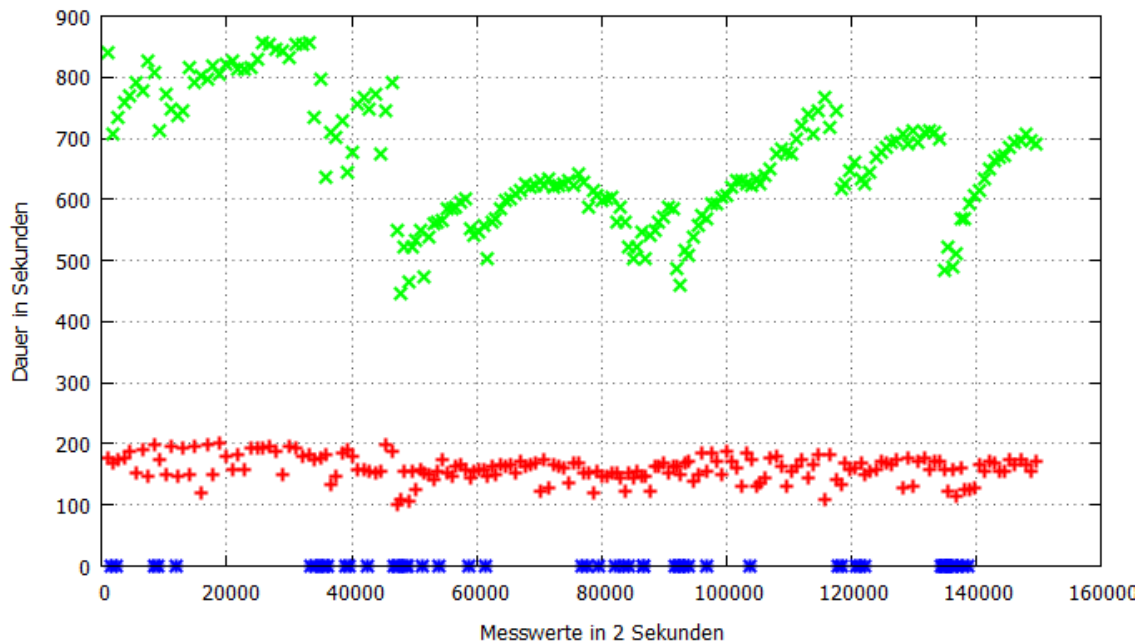


Abbildung 12: Dauer der Kühlphasen (rot), Pausen dazwischen (grün) und Indikator für die Kühlschranklampe (blau) mit Interaktionen

Zu bemerken ist hier, dass die Messungen in diesem Beispiel mit einem anderen Kühlschrankmodell (Kühlschrank B) durchgeführt wurden. Auffällig ist zunächst vor allem die allgemein deutlich kürzere Dauer der Kühlphasen und daraus resultierend auch der Pausen zwischen zwei solcher Intervalle. Während das Gerät aus den bisherigen Betrachtungen grob alle zwei Stunden für einen Zeitraum von 45 Minuten kühlt, lässt sich hier nun grob eine 5-minütige Kühlphase alle 30 Minuten feststellen.

Dass das bisher verwendete Vorgehen allerdings nicht zuverlässig anwendbar ist, lässt sich bereits ohne weitere Analyse feststellen, wenn man die durchschnittliche Kühldauer über die Messreihe mit hoher Aktivität und der ohne jegliche Aktivität vergleicht. Letztere ist mit 368 Sekunden sogar höher als die der Messung um die Weihnachtstage (320 Sekunden). Allein diese Tatsache lässt es äußerst fragwürdig erscheinen, dass bei diesem Kühlschrankmodell die Kühldauer höher wird, wenn die Türe geöffnet wird. Anschaulich lässt sich das etwa in Abbildung 12 darstellen. Die roten Messpunkte stellen dort die Dauer der einzelnen Phasen dar, die keine Korrelation in vermuteter Form mit den blauen Indikatoren für den Stromverbrauch der Lampen erkennen lassen. Nun erscheint es aber äußerst fragwürdig, dass das Erwärmen der Luft durch das Öffnen der Türe keinerlei Auswirkung auf das Kühlverhalten des Geräts hat, vor allem da auch kein höherer Stromverbrauch während eines Intervalls und damit beispielsweise eine höhere Kühlleistung in glei-

cher Zeit nachgewiesen werden kann.

Wenn man nun allerdings nicht die Dauer einer Kühlphase, also das Intervall [Start der Phase, Ende der Phase] betrachtet, sondern die Pause zwischen der letzten und der aktuellen Phase [Ende der letzten Kühlphase, Start der aktuellen Kühlphase] betrachtet, lassen sich über diese Werte doch Aussagen über das Verhaltensmuster des Benutzers treffen. Die Länge dieser Pausenintervalle sind nun in Abbildungen 11 und 12 als grüne Kreuze dargestellt. Wenn man diese nun zunächst in Abbildung 11 betrachtet, fällt auf, dass sich diese relativ konstant im Rahmen von 900 bis 1000 Messpunkten bewegen. Als Vergleich lassen sich im nächsten Schritt die Werte der Messung während hoher Aktivität aus Abbildung 12 heranziehen. In diesem Fall ist erstens die Spanne, in der sich die Werte bewegen um einiges größer (zwischen 450 und knapp 900), was auf jeden Fall die Grundannahme, dass während einer längeren Phase gleichbleibender Bedingungen eine geringere Streuung der Messwerte auftritt, bestätigt. Zweitens lässt sich auch eine deutliche Korrelation zwischen den Abständen zwischen zwei Kühlphasen des Geräts und den Lampenevents feststellen. Letztere sind in Abbildung 12 als blaue Kreuze auf der x-Achse dargestellt. Auch hier lassen sich, wie schon im vorherigen Beispiel, gewisse Cluster feststellen, die sich, bei genauer Betrachtung um übliche Essenszeiten bewegen. Eine weitere Auffälligkeit des Verlaufs in Abbildung 12 sind die verschiedenen "Grundlevel" der Pausendauern zwischen dem Beginn der Betrachtung bis etwa 45000, dann allgemein deutlich niedrigere Werte bis etwa 105000 und danach wieder tendenziell höhere Intervalle. Ein möglicher Erklärungsversuch geht nun über zwei Faktoren. Zum einen über eine gleichzeitige, grobe Temperaturmessung¹ der Zimmertemperatur, in dem sich der Kühlschrank befindet, sowie über die Erklärung der sonstigen Einflüsse. Der Messwert um etwa 45000 entspricht 16:30 Uhr an Heiligabend. Während sich die Zimmertemperatur davor zwischen 16 und 17 Grad bewegte, wird der Raum nun in der Phase der Essensvorbereitung beheizt und hat etwa 22 Grad. Der Zeitraum zwischen 62000 und 75000 entspricht der Zeit zwischen 1:00 und 9:00 ohne irgendwelche Aktivität, die Temperatur im Raum bleibt aber relativ konstant auf über 20 Grad. In diesem Zeitraum geht die Dauer der Pausen wieder deutlich nach oben, was auf eine geringere benötigte Kühlung hinweist und sich mit den bisherigen Annahmen deckt. Der folgende Tag (erster Weihnachtstag) zwischen 75000 und 105000 Messpunkten (anhand der Timestamps entspricht dies etwa dem Zeitraum zwischen 9:00 Uhr und 1:00 Uhr) lässt wieder das zu erwartende Muster von kürzeren Kühlpausen nach Öffnen der Türe, mit folgenden größer werdenden Intervallen beobachten. Der Temperaturverlauf an diesem Tag beginnt grob mit einem konstanten Wert von um die 20-22 Grad und sinkt dann ab dem Abend wieder auf einen unge-

heizten Wert von etwa 17 Grad ab. Dieses Absinken der Temperatur lässt sich nach den Türevents ab etwa 105000 (also 1:00 Uhr) auch in Abbildung 12 nachvollziehen, da sich die Pausenintervalle dann wieder, im Gegensatz zu der Nachtphase den vorherigen Tages in noch geheiztem Zustand, in Richtung 800 bewegen.

Aus diesen Überlegungen folgt nun eine Erweiterung der bisher verwendeten Analyse um die Pause zwischen zwei Kühlphasen. Da die Daten bereits vorliegen ist es ein einfaches, diesen zusätzlichen Wert zu generieren. Konkret geht es um den Zeitraum zwischen dem Ende der letzten Kühlphase und dem Beginn der Aktuellen. Die dem zugrunde liegende Logik wäre also, dass das Aufwärmen des Innentemperatur des Kühlschranks hier zu einer kürzeren Pause führt, wenn dieses während der letzten Kühlphase bzw. während der gerade betrachteten Pause erfolgt ist. Aus der Betrachtung ist jedoch auch deutlich geworden, dass sich damit vermutlich nur marginal bessere Ergebnisse erzielen lassen, da erstens die Phasen bei diesem Kühlschrankmodell sehr kurz sind und die Einflüsse der Lampenevents nicht unmittelbar sichtbar sind, sondern eher im Laufe von mehreren Zyklen Wirkung zeigen. Letzterer Sachverhalt lässt sich etwa in Abbildung 12 während den Nachtphasen (62000-75000 und 105000-118000) gut sehen. Während diesen Zeiträumen tritt keine Interaktion mehr von Nutzerseite auf, die Pausenintervalle nehmen aber über einen Zeitraum von mehrere Stunden langsam zu und passen sich nicht etwa sofort an einen Wert an.

Es erscheint also sinnvoll, nicht nur jeweils eine Pause zwischen zwei Aktivitätsphasen des Kühlschranks zu betrachten, sondern dies auf mehrere Zyklen auszudehnen. Damit soll eine relativ zuverlässige Methode erreicht werden, mit der sich etwa über einen Zeitraum von zwei Stunden eine Aussage darüber treffen lässt, ob ein Benutzer anwesend war oder nicht. In dieser Betrachtung müssen jedoch auch die unterschiedlichen "Grundlevel" beachtet werden, als deren Ursache unter anderem die Raumtemperatur vermutet wird.

4.2.4 Nutzerverhalten über mehrere Kühlphasen

Wenn man noch einmal die Werte aus Abbildung 12 betrachtet, ergibt sich über die gesamte Messdauer hinweg eine durchschnittliche Pausendauern zwischen zwei Kühlphasen von 654 Messwerten. Analog zu dem Vorgehen bei den Dauern der Kühlphasen wäre es nun eine Möglichkeit, entweder ab einer gewissen Abweichung vom Durchschnitt oder wieder über eine komplementäre Verteilungsfunktion einen Wert zu bestimmen, unter dem man eine Einwirkung des Nutzer vermuten würde. Hier fällt aber direkt ein fundamentales Problem an dieser Vorgehensweise im Zu-

¹Das Temperaturprotokoll befindet sich als Anhang auf der CD

sammenhang mit der derzeitigen Betrachtung der Pausen in Abbildung 12 auf. Die Intervalle der Pausen sind nicht nur vom Öffnen der Türe abhängig, sondern augenscheinlich auch von anderen Faktoren, etwa der Umgebungstemperatur. Somit stößt man mit dieser Vorgehensweise recht schnell an Grenzen. Wenn man beispielsweise die Werte während Aktivität des Nutzers, erkannt anhand des Stromverbrauchs der Kühlschranklampe, bei etwa 135000 mit denen während einer ruhigen Nachtphase bei 70000 vergleicht, sind diese Werte fast identisch. Es ist also naheliegend nicht den gesamten Trace zu wählen, sondern die Vergleiche nur auf eine lokale Umgebung zu beziehen.

Im folgenden Vorgehen wird die Gesamtmessung in Abschnitte von jeweils fünf Kühlphasen- und deren Pausen unterteilt. Wie bereits erwähnt, bewegt sich ein solcher Zyklus etwa im Bereich von 850 Messpunkten, also einer knappen halben Stunde. Die Betrachtung von fünf aufeinanderfolgenden Zyklen entspricht damit also einer Auflösung von etwa zwei Stunden. Es wird nun ein Öffnen der Kühlschranktür des Benutzers erwartet, wenn sich die durchschnittliche Pausendauer eines solchen Quintetts im Vergleich zu der durchschnittlichen Pausendauer der vorherigen Fünfergruppe verringert hat (bzw. in diesem Beispiel nicht erhöht hat, also die Änderung 0 oder weniger betrifft). Die Logik dahinter entspricht also der Feststellung, dass sich nach einem Lampenevent (blau in Abbildung 12) die Dauer der Kühlphasen schlagartig nach unten bewegt und die Steuerung des Kühlschranks diese dann langsam wieder nach oben korrigiert. Während diesen Phasen gibt es zwar einige Ausreißer, die dem angenommenen Muster nicht oder nur grob folgen, dies fällt aber bei der Betrachtung des Durchschnitts über mehrere Intervalle nicht weiter ins Gewicht.

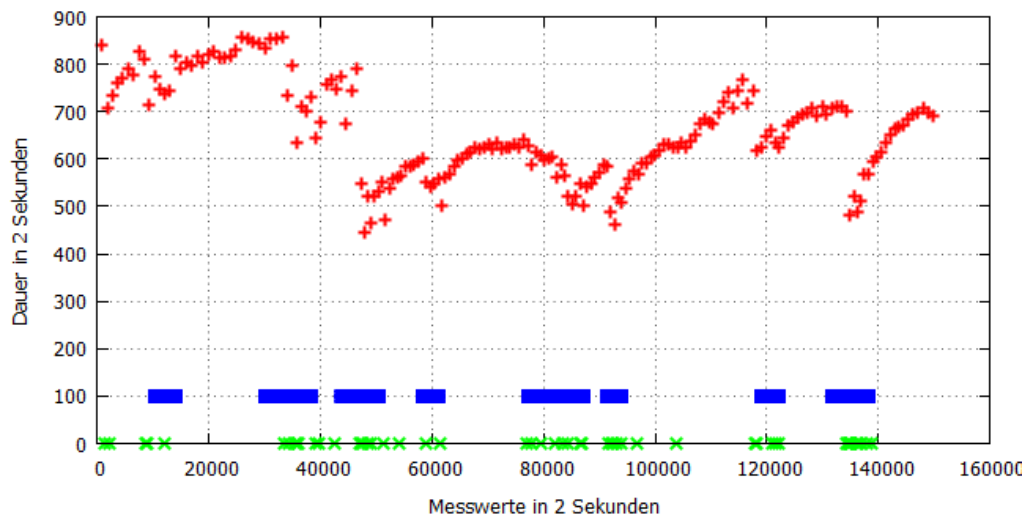


Abbildung 13: Dauer der Pausen (rot), Lampenevents (grün) und erkannte Intervalle der Anwesenheit (blau)

Abbildung 13 stellt nun eine Auswertung nach diesem Muster dar. Während die Pausendauern zwischen den einzelnen Kühlphasen, sowie die Lampenevents, auf der x-Achse aufgetragen, bereits bekannt sind, verdeutlichen die blauen Bereiche nun die Intervalle, die nach den im letzten Abschnitt erläuterten Bedingungen eine im Durchschnitt sinkende Pausendauer und damit vermutete Interaktion des Benutzers haben. Auch wenn mit diesem Verfahren nicht alle Events erkannt werden und vor allem einzelne Lampenevents außerhalb der erkannten Bereiche liegen (beispielsweise das Lampenevent bei etwa 105000), ist das Resultat als Ganzes bereits zufriedenstellend. Die Nachtphasen werden gut erkannt und um Cluster von grünen Kreuzen ist die Erkennung zuverlässig.

Es bietet sich an, die Auswertung analog zum Vorgehen bei der Dauer der Intervalle zu quantifizieren.

true positive t_p entspricht 5er-Zyklen, die eine abfallende Pausendauer haben und in denen ein Lampenevent auftritt

true negative t_n entspricht 5er-Zyklen ohne abfallende Pausendauer und ohne Lampenevents

false positive f_p entspricht 5er-Zyklen mit abfallender Pausendauer, aber ohne registrierten Lampen

false negative f_n entspricht 5er-Zyklen ohne abfallender Pausendauer, aber mit Lampenevents

Zusätzlich wird eine Betrachtung mehrerer Grenzwerte vorgenommen, wobei sich

dieser Grenzwert nun auf die Höhe der Absenkung der durchschnittlichen Pausendauer zur letzten 5er-Gruppe bezieht. Während in Abbildung 13 dieser Grenzwert 0 beträgt, also schlichtweg jegliche Verringerung der durchschnittlichen Dauer der Pausen in diesem Intervall positive zurückliefert, sollen nun also eine Reihe von verschiedenen Werten, analog zu Tabelle 2, betrachtet werden. Dies ist wichtig, da das Verfahren ja auch bei der Betrachtung des Zeitraums aus Abbildung 11 funktionieren soll, in dem offensichtlich eine natürliche Abweichung auch ohne Interaktion präsent ist.

Grenzwert	t_p	t_n	f_p	f_n	$P(\text{richtigKlassifiziert})$
0	13	14	0	8	0.77
-5	13	14	0	8	0.77
-10	13	14	0	8	0.77
-15	12	14	0	9	0.74
-20	11	14	0	10	0.71
-25	7	14	0	14	0.6
-30	7	14	0	14	0.6

Tabelle 3: Ergebnis verschiedener Grenzwerte (bei Pausendauern vgl. Abb. 12)

Grenzwert	t_p	t_n	f_p	f_n	$P(\text{richtigKlassifiziert})$
0	0	9	12	0	0.43
-5	0	12	9	0	0.57
-10	0	17	4	0	0.80
-15	0	21	0	0	1.0
-20	0	21	0	0	1.0

Tabelle 4: Ergebnis verschiedener Grenzwerte (bei Pausendauern vgl. Abb. 11)

Das Ergebnis spiegelt die zu erwartende Tendenz wieder. In Falle des Zeitraums um die Weihnachtstage mit viel Interaktion sinkt die Erkennungsrate mit Hochschrauben des Grenzwertes ab, während man bei den Tagen ohne Interaktion eine gewisse Schwelle benötigt, um sinnvolle Ergebnisse zu erhalten, da sonst die natürliche Abweichung der Pausendauern oft zu einer fälschlicherweise positiven Erkennung (f_p) führt. Ab einem Grenzwert von -15 im Durchschnitt (das bedeutet, dass die Erkennung nur triggert, wenn die durchschnittliche Dauer über die 5er-Gruppen um mehr als 15 Messpunkte oder 30 Sekunden gesunken ist) funktioniert die Erkennung in letzterem Fall komplett korrekt. Laut Tabelle 3 lässt sich dieser Wert auch hier mit zufriedenstellender Genauigkeit anwenden und letztlich sorgt das Anheben der Schwelle um 15 Sekunden nur in einem Fall zu einer neuen falschen Erkennung.

4.2.5 Anwendung der Pausenbetrachtung auf andere Modelle

Abschließend scheint es noch interessant den Kühlschrank aus Kapitel 4.2.2 in Hinblick auf die Pausendauern zu untersuchen. Eine Betrachtung über mehrere Zyklen hinweg, wie es nun bei dem anderen Gerät erfolgt ist, erscheint durch die relativ geringe Anzahl an Kühlphasen nicht angebracht. Dies hat zum Einen mit der kürzeren Messdauer zu tun, zum anderen aber natürlich auch mit der bereits erwähnten unterschiedlichen Länge der Zyklen von Gerät zu Gerät.

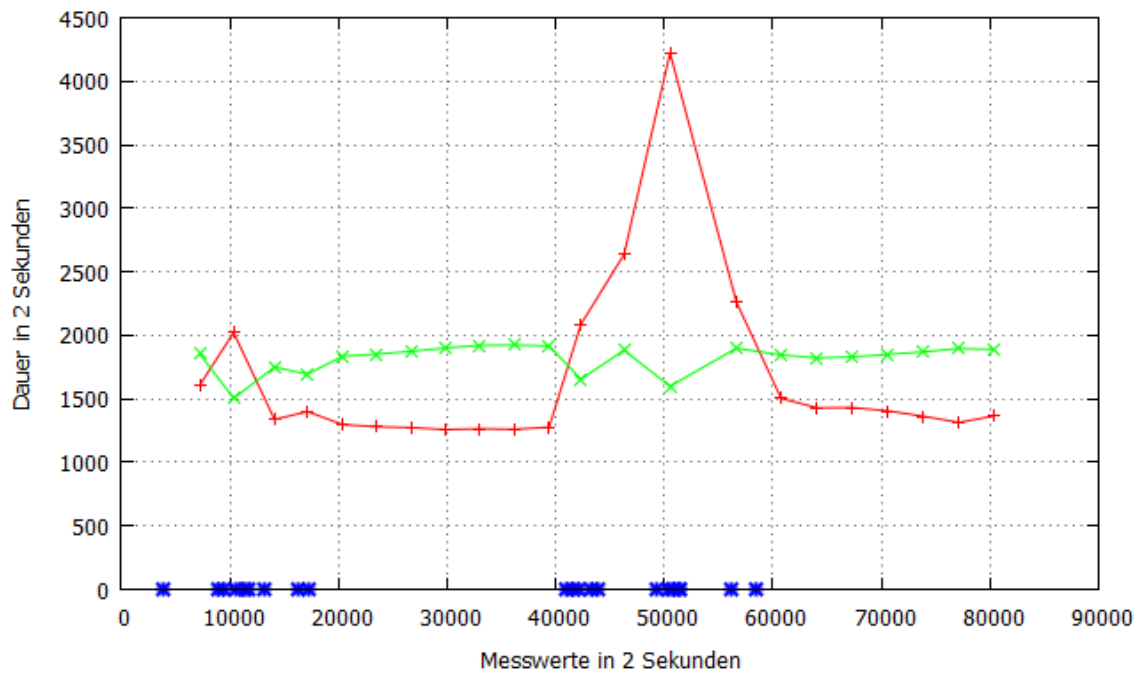


Abbildung 14: Dauer der Kühlphasen (rot), Dauer der Pausen (grün) und erkannte Lampen (blau)

Trotzdem sieht man in Abbildung 14 eine Bestätigung der bisher gefundenen Zusammenhänge. Die durch die grüne Kurve dargestellte Länge der einzelnen Pausen zwischen den Kühlphasen bzw. in diesem Fall der Pause seit der letzten Kühlphase verringert sich immer wenn eine Häufung von Lampenevents nachgewiesen werden konnte. Des Weiteren sieht man auch eine entgegengesetzte Tendenz der angesprochenen Pausendauer zu der in rot dargestellten Dauer der Kühlphasen. Die anfangs beobachtete Korrelation zwischen Öffnen der Türe und der Länge der folgenden Kühlphase lässt sich nun also durch die Verringerung der Pause bis zu dieser Kühlphase erweitern. Letzterer Zusammenhang wurde auch bei den zweiten Gerät deutlich, aber hier konnte keine signifikante Aussage über die Kühldauer getroffen werden. Bei der Messung von Kühlschranks B zeigte sich, etwa in Abbildung 13, ein langsames Annähern der Pausendauer an einen ggf. von der Umgebungstemperatur abhängigen "Normalwert" nach dem signifikanten Abfall dieser Dauer nach einem Öffnen der Kühlschranktür. Dieses Verhalten lässt sich in Abbildung 14 auch Kühlschrank A attestieren. In den Nachtphasen zwischen rund 20000 und 40000, sowie ab 60000 Messpunkten stieg die Dauer der Pausen vor den einzelnen Kühlphasen jeweils leicht an. Die Dauer der Kühlphasen weisen entsprechend einen leicht inversen Trend auf.

4.2.6 Zusammenhang mit dem Energieverbrauch

Ein bisher gar nicht betrachteter Wert der Aufzeichnung des Raspberry Pi ist die Gesamtenergie des Smart Meter. Zunächst soll theoretisch untersucht werden, welche Beobachtungen hierbei zu erwarten sind. Der Anstieg der Energie kann bei einer isolierten Betrachtung eines Kühlschranks als Kombination der Kühl- und Pausendauern gesehen werden. Da sich die Höhe des Startpeaks, der folgende Abfall und schlussendlich der Endabfall immer etwa gleich verhalten, ist zu erwarten, dass eine längere Kühlphase auch mehr Energieverbrauch als Folge hat. Gleichzeitig ist es offensichtlich, dass auch eine geringere Pausendauer eine Erhöhung des Energieverbrauchs in gleichbleibenden Zeitschritten erzeugt. Unter der Bedingung, dass die erste Annahme korrekt ist, muss über diesen Zusammenhang zwischen Pausendauer und Energie nicht weiter diskutiert werden. Es ist also sinnvoll, zunächst zu bestätigen, dass eine längere Kühlphase direkt einen höheren Energieverbrauch in dieser Zeit bedingt. Die Lampenevents spielen hier zwar auch etwas hinein und müssten streng genommen in dieser Betrachtung nicht gemessen werden, der Einfachheit halber soll dies aber vernachlässigt werden.

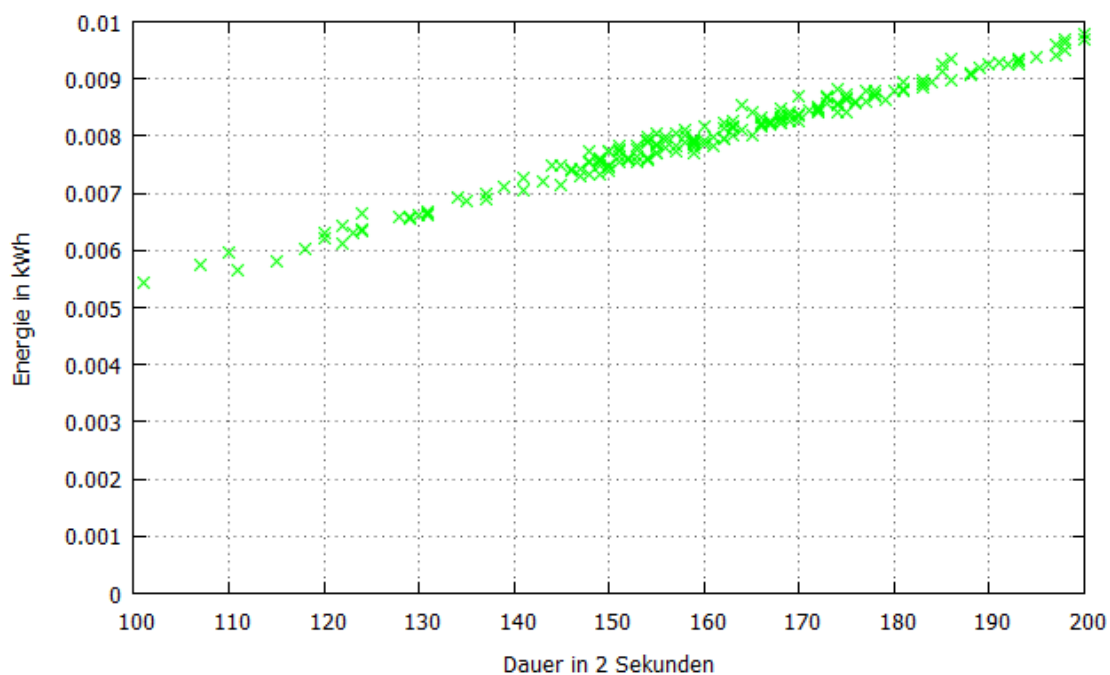


Abbildung 15: Energieverbrauch bei versch. Kühldauern bei Kühlschrank B

Abbildung 15 zeigt diesen Zusammenhang deutlich und lässt einen grundsätzlich linearen Anstieg des Energieverbrauchs erkennen. Dies deckt sich mit den getroffenen theoretischen Annahmen. Ein ähnlicher Verlauf zeigt sich auch bei Kühlschrank A.

Im nächsten Schritt ist es nun möglich, die bisherigen Ergebnisse auch anhand der Energie nachzuvollziehen.

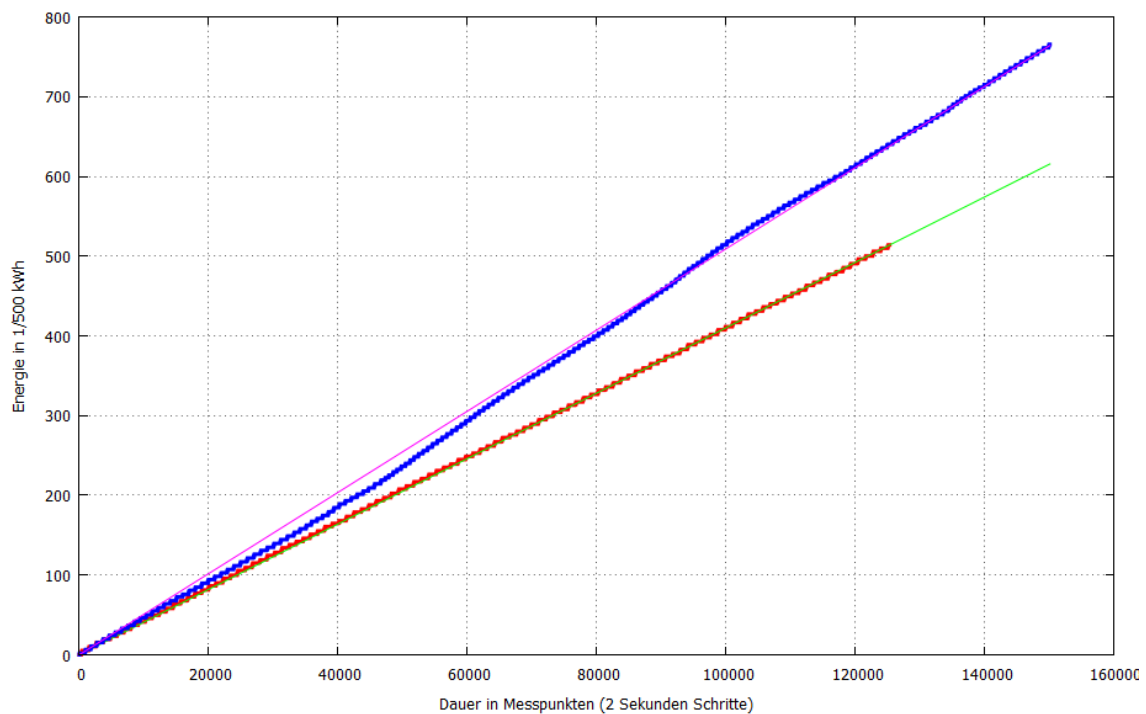


Abbildung 16: Energieverbrauch bei Aktivität und Inaktivität im Vergleich

Zunächst zur Erklärung der in Abbildung 16 dargestellten Daten. Die rote Treppenkurve stellt den gesamten gemessenen Energieverbrauch während der Phase ohne Interaktion des Nutzers dar. Dieser beträgt in diesem Fall (die Daten sind um den Faktor 500 skaliert) also im Zeitraum der Messung von rund 125000 Zeitschritten oder 70 Stunden knapp über 1 kWh. Die blaue Kurve stellt den selben Sachverhalt anhand der Messdaten während der Phase mit Interaktion dar. Die verwendeten Daten sind bereits aus Abbildungen 11 bzw. 12 bekannt. Die beiden Hilfslinien (grün bzw. lila) dienen nur zur Veranschaulichung und sollen es erleichtern, Abweichungen nach unten bzw. oben zu identifizieren. Da Kühltank B keine nachvollziehbare Änderung der Kühltankdauern aufweist, ist die Änderung im Energieverbrauch primär durch die schnellere Abfolge der Kühltankphasen bedingt. Der in Abbildung 11 vergleichsweise konstante Verlauf dieser Pausen ist nun auch in Abbildung 16 deutlich erkennbar und weicht nahezu nicht von der grünen "Mittellinie" ab. Wenn man im Vergleich dazu die Messwerte der Aktivitätsphase (blau) betrachtet, fällt zunächst der höhere Energie insgesamt auf; im gleichen Zeitraum fällt dieser mit rund 1,25 kWh um ein Viertel höher aus. Auch bei genauerer Betrachtung des Verlaufs lassen sich unterschiedliche Steigungen erkennen. Die Dauer der Pausen lässt sich in

Abbildung 12 grob in drei Bereiche unterteilen. Zunächst bis rund 40000 Messpunkte eher hohe Pausen, dann bis 100000 Messpunkte kürzere Pausen und am Ende wieder leicht höhere Pausendauern. Diese Bereiche lassen sich auch in Abbildung 16 wiederfinden. Zunächst hängt der Energieverbrauch eher nach unten durch, was mit der beobachteten eher längeren Pausendauer übereinstimmt. Im Mittelbereich ist eine größere Steigung zu sehen und wieder ab etwa 100000 Messpunkten knickt diese Steigung wieder leicht nach unten ab. Zusätzlich lassen sich die größeren Ansammlungen an Lampenevents auch in diesem groben Maßstab noch ansatzweise erkennen. Eines dieser Cluster befindet sich nach 80000, was auch im Verlauf der blauen Kurve eine größere Steigung bewirkt. Bei rund 120000 lässt sich ein weiteres Cluster feststellen, was sich auch wieder in einer leichten Korrektur nach oben bemerkbar macht.

Zu bemerken ist allerdings, dass dieses Verfahren in einer überlagerten Kurve nicht praktikabel ist und nur bei einer isolierten Betrachtung des Kühlschranks funktioniert. Mit anderen Geräten als Störfaktoren könnte man lediglich eine Annäherung, etwa über eine exponentielle Funktion, zwischen den Start- und Endevents den Energieverbrauch simulieren, was aber vermutlich nicht zielführend ist. Trotzdem bestätigt die Untersuchung auch über diesen Weg einen Zusammenhang zwischen Nutzerinteraktion mit dem Kühlschrank und dem Stromverbrauch von diesem.

5 Fazit

5.1 Zusammenfassung und Bewertung

Das Vorgehen zeigt deutlich, dass sich aus den Messwerten des Smart Meter in diesem Experiment klare Rückschlüsse auf die Anwesenheit des Nutzers ziehen lassen. Im Falle des isolierten Vorliegens des Kühlschranksverbrauchs ist es sehr einfach, die entsprechenden Events herauszulesen und auch bei der Überlagerung mehrerer Geräte funktioniert dies noch zufriedenstellend. In diesem Zusammenhang ist es auch wichtig zu bemerken, dass über den Tag hinweg wenig "außergewöhnliche Peaks" in der überlagerten Kurve eines Haushalts auftreten sollten. Die meisten konstanten Verbraucher, also etwa Lampen, haben einen recht einfach herauszufilternden Stromverbrauch. Das Hauptproblem bei der Erkennung der einzelnen Zustände bzw. Übergänge zwischen diesen, stellen andere Verbraucher dar, die genau im Zeitfenster der 6-10 Sekunden des Startevents bzw. Endevents einen höheren Peak haben. Hier muss man sich ab einem gewissen Punkt mit falschen Erkennungen abfinden und es besteht lediglich die Option, diese so gut es geht abzufangen bzw. in der Verarbeitung zu berücksichtigen. Bei der Entfernung des Kühlschranks ist das erzielte Ergebnis bei dem primären Versuchsobjekt, Kühlschrank A, gut und das verbleibende Rauschen sollte bei der Größenordnung, in der die Erkennung funktionieren soll, keine weitere Rolle mehr spielen.

Was die Aussagekraft in Hinblick auf datenschutzrelevante Spuren betrifft, lässt sich anhand eines Kühlschranks eine gute Quote erreichen und somit lässt sich auch die ursprüngliche Annahme, dass man aus den Stromverbrauchsdaten klare Schlüsse auf das Verhalten der Nutzer ziehen kann, bestätigen. Es reicht hier bereits aus, die unterschiedlichen Dauern aus Abbildungen 11 und 12 miteinander zu vergleichen. Bereits ohne die Ergebnisse weiter zu interpretieren, wird allein durch die Struktur der Pausendauern deutlich, dass in Abbildung 12 eine gewisse Aktivität vorlag. Wenn man die Auswertungen gemäß den Tabellen 1,2 und 3 hinzunimmt, kann diese Struktur auch quantifiziert werden. Es ist vor allem möglich die Nachtphasen bzw. Phasen von Inaktivität zu identifizieren und zum anderen lassen sich die Cluster um Essenszeitpunkte herum, also mehrere Öffnungen des Kühlschranks innerhalb eines kurzen Zeitintervalls, auch im Stromverbrauch sehen. Gut deutlich wird dies etwa in Abbildung 14 bei Kühlschrank A. Auch bei Gerät B ist eine Korrelation der Lampenevents mit der blau markierten erkannten 5er-Gruppen sichtbar. Gerade bei Kühlschrankmodellen mit kurzen Kühlphasen scheint der Weg über eine Betrachtung von mehreren Kühlphasen sinnvoll zu sein. Hier stellt sich natürlich auch die Frage, für welchen Verwendungszweck die Messdaten und deren Ergebnisse

verwendet oder missbraucht werden sollen.

Durch die Wahl des Kühlschranks als primäres Untersuchungsobjekt, hat man durch die Lampenevents einen großen Vorteil gegenüber anderen Geräten gesehen. Es war hier nicht notwendig ein ausführliches Protokoll über die Aktivität zu führen, um die Ergebnisse bewerten zu können, sondern die Messung gibt das Werkzeug dafür quasi inhärent vor.

Die Messgenauigkeit von einem Messwert alle 2 Sekunden ist vergleichsweise hoch, wenn man die Start- und Endevents aber auch bei geringerer Auflösung noch erkennen kann, funktioniert das Vorgehen genauso.

5.2 Ausblick

In Hinblick auf den Kühlschrank wäre es in erster Linie interessant weitere Modelle zu überprüfen und dabei vor allem die Aussagekraft der beiden Merkmale Kühldauer und Pausendauer in Zusammenhang mit den nachgewiesenen Türöffnungen zu untersuchen. Die beiden in dieser Arbeit untersuchten Kühlschränke zeigten beide eine klare Korrelation der Nutzerinteraktion mit der Pausendauer, bei Kühlschrank A lies sich auch über die Kühldauer Rückschlüsse ziehen. Es wäre also interessant zu sehen, wie sich dies bei weiteren Geräten verhält und ob man evt. die beiden Werte so zusammenfassen kann, dass man bei allen Modellen diese Schlüsse ziehen kann. Auch die Erkennung der Signatur bzw. das Entfernen dieser kann noch weiter untersucht werden. Die Abfallphase der Leistung ist zwar sowohl bei Kühlschrank A, als auch bei Kühlschrank B grob exponentiell, bei letzterem Gerät ist dies aber nicht so sauber ausgeprägt. Eventuell wäre es angebracht bei anderen Kühlschrankmodellen hier ein angepasstes Beschreibungsmodell zu verwenden. Schlussendlich birgt auch der Zusammenhang der Messwerte mit Außentemperatur und Inhalt des Kühlschranks noch Untersuchungspotential. Die gewonnenen Daten legen die Vermutung nahe, dass beide Faktoren Einfluss auf die Dauern bzw. Pausen haben, dies müsste aber, um definitive Aussagen treffen zu können, noch weiter untersucht werden. Mit diesen Erkenntnissen wäre es dann denkbar, die Erkennung der Anwesenheit noch von weiteren Variablen abhängig zu machen und diese nicht nur auf die Lampenevents zu stützen.

Die gewonnenen Erfahrungen anhand des Kühlschranks können nun in ein allgemeines Verfahren umgesetzt werden, mit dem auch weitere Geräte analysiert werden können. Im Wesentlichen lässt sich das Vorgehen aus den Kapiteln 3 und 4 in folgende Unterpunkte zusammenfassen:

Modellieren eines Gerätes Die meisten Verbrauchsgeräte in einem Haushalt weisen eine eindeutige oder zumindest gut erkennbare Signatur im Stromverbrauch auf. Der erste Schritt besteht also darin, diese Signatur zu modellieren. Hierfür müssen Gemeinsamkeiten zwischen mehreren Vorkommen des gleichen Geräts gefunden und so implementiert werden, dass möglichst wenig Fehlererkennung stattfindet. Im Optimalfall sollte dieses Vorgehen auch auf verschiedene Modelle des Geräts anwendbar sein.

Erkennen der Geräte und Weiterverarbeitung Der in Schritt 1 formulierte Algorithmus kann nun verwendet werden, um das Vorkommen des zu untersuchenden Geräts aus dem Stromverbrauch zu extrahieren. Hier kommen, in Anlehnung an den Kühlschrank, mehrere relevante Punkte in Frage. Auf einfachster Ebene ist natürlich das reine Vorkommen, also der Zeitpunkt der Nutzung des Geräts, interessant. Bei unterschiedlicher Länge der Nutzung kann dann auch, analog zu den Kühldauern und Pausendauern, die Länge eines Vorkommens bzw. die Pause zwischen zwei Vorkommen in der Stromverbrauchskurve interessant sein. Zusätzlich kann, ähnlich wie in [GJL12], auch die genaue Höhe des Verbrauchs eines Geräts Aufschluss auf das Nutzerverhalten geben. In diesem Schritt sollen diese Zeitpunkte, Dauern oder andere Werte möglichst zuverlässig gefunden und zur Weiterverarbeitung aufbereitet werden. Teil dieser Verarbeitung kann auch das Entfernen des Geräts aus der überlagerten Kurve sein, um so eine Reihe von verschiedenen Verbrauchern nacheinander zu identifizieren.

Korrelation zwischen den Daten und Nutzerverhalten finden Im zweiten Schritt geht es dann darum anhand der erstellten Datensammlung Rückschlüsse auf das Verhalten des Nutzers zu ziehen. Auch hier ist es wieder wichtig den Kontext des zu untersuchenden Geräts zu betrachten und die Art der möglichen Schlüsse festzulegen. Während etwa beim Kühlschrank recht komplexe Betrachtungen erfolgten, wären die Schlüsse bei einem Gerät, das nur durch das Vorkommen die Anwesenheit eines Benutzers vermuten lässt und bei dem die Dauer der Nutzung keine weitere Interpretation erlaubt, schneller erschöpft.

Das beschriebene Vorgehen kann also als Art Blaupause verwendet werden, um weitere Geräte zu analysieren. Am Interessantesten erscheinen hier Geräte, die die in Kapitel 3.3 vorgestellten Kriterien möglichst gut erfüllen und dann komplexere Aussagen über die Privatsphäre erlauben.

6 Quellenverzeichnis

- [Bun14] BUNDESAMT FÜR SICHERHEIT IN DER INFORMATIONSTECHNIK: *Das Smart Meter Gateway, Sicherheit für intelligente Netze*. 2014. – <https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/-Broschueren/SmartMeter/Broschuere-SmartMeter.pdf>, abgerufen am 16.02.2014
- [Ern13] ERNST&YOUNG GMBH: *Kosten-Nutzen-Analyse für einen flächendeckenden Einsatz intelligenter Zähler*. 2013. – <http://www.bmwi.de/DE/Mediathek/publikationen,did=586064.html>, abgerufen am 16.02.2014 (angefügt als Datei ernstyoung.pdf)
- [GJL12] GREVELER, Ulrich ; JUSTUS, Benjamin ; LÖHR, Dennis: Identifikation von Videoinhalten über granulare Stromverbrauchsdaten / Fachhochschule Münster. 2012. – Forschungsbericht. – http://1lab.de/pub/GrJuLo_Smartmeter.pdf, abgefrufen am 16.02.2014 (angefügt als Datei GrJuLoSmartmeter.pdf)
- [Kon12] KONFERENZ DER DATENSCHUTZBEAUFTRAGTEN DES BUNDES UND DER LÄNDER UND DÜSSELDORFER KREIS : *Orientierungshilfe datenschutzgerechtes Smart Metering*. 2012. – http://www.bfdi.bund.de/SharedDocs/Publikationen/Entschliessungssammlung/DSBundLaender/Orientierungshilfe_SmartMeter.html, abgerufen am 16.02.2014 (angefügt als Datei Orientierungshilfe.pdf)
- [Mü10] MÜLLER, Klaus J.: Gewinnung von Verhaltensprofilen am intelligenten Stromzähler. In: *Datenschutz und Datensicherheit Ausgabe 6, 2010* (2010)
- [PS13] PETRUSCH, Luca ; SCHULTHESS, Christoph: *Smart-Metering with the Raspberry Pi*. 2013. – Hochschule der Medien, Stuttgart - Projektdokumentation